DOCUMENT RESUME

ED 157 710 SE 024 514

AUTHOR	Simon, Charles W.
TITLE	Design, Analysis, and Interpretation of Screening Studies for Human Factors Engineering Research. Revised Edition.
INSTITUTION	Canyon Research Group, Inc., Westlake Village, Calif.
SPONS AGENCY	Air Force Office of Scientific Research, Washington, D.C.
REPORT NO	TR-CWS-03-77A
PUB DATE	Jun 78
CONTRACT	F44620-76-C-0008
NOTE	251p.; Update to ED 150 005; Contains marginal legibility in Appendices; Pages 194 and 198 removed due to oversize and marginal legibility
EDRS PRICE	MF-$0.83 HC-$14.05 Plus Postage.
DESCRIPTORS	*Behavioral Science Research; Factor Analysis; *Human Engineering; *Research Design; Scientific Methodology; *Screening Tests; *Statistical Analysis; Trend Analysis

ABSTRACT

Methods are described for constructing Resolution IV Screening designs that are robust to linear, quadratic, and cubic trend effects and will also keep the number of factor-level changes for some variables at a minimum. Complete designs, capable of screening up to 32 variables, are provided along with appropriate methods of analysis. Numerous criteria for selecting non-trivial variables are discussed, including: mean difference, eta squared, cumulative probability, and half-normal plots. How to handle subjects in an experiment is described when their characteristics can be included as experimental factors and when they cannot be, and when subjects are merely a form of replication. The values of introducing multiple center points into screening designs are discussed. Lack-of-fit tests are provided to help decide whether a second- or third-order response surface is needed. Numerous methods of analyzing screening studies with multiple responses are described. A method is given for developing a prediction equation with data collected from an incomplete screening design. (Author/HM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157710

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DESIGN, ANALYSIS, AND INTERPRETATION OF SCREENING STUDIES FOR HUMAN FACTORS ENGINEERING RESEARCH

Charles W. Simon

Technical Report CWS-03-77A

Research sponsored by the Air Force Office of Scientific Research (AFSC), United States Air Force, under Contract No. F44620-76-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

September 1977
Revised June, 1978

Canyon Research Group, Inc.
741 Lakefield Road, Suite B
Westlake Village, California, 91361

SE 024 514

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CWS-03-77A	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Design, Analysis, and Interpretation of Screening Studies for Human Factors Engineering Research		5. TYPE OF REPORT & PERIOD COVERED Technical Report Sept. 1976 - Sept. 1977
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Charles W. Simon		8. CONTRACT OR GRANT NUMBER(s) F44620-76-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Canyon Research Group, Inc. 741 Lakefield Road, Suite B Westlake Village, CA 91361		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling Air Force Base Washington, D. C., 20332		12. REPORT DATE September 1977
		13. NUMBER OF PAGES 220
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution list included at end of report.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

CWS-03-77, revised June 1978 from original document, becomes CWS-03-77A.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Screening designs
Fractional factorials
Trend-robust designs
Multiple-response analysis
Half-normal plots
Lack of fit analysis
Center points

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Methods are described for constructing Resolution IV screening designs that are robust to linear, quadratic, and cubic trend effects and will also keep the number of factor-level changes for some variables at a minimum. Complete designs, capable of screening up to 32 variables, are provided along with appropriate methods of analysis. Numerous criteria for

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

selecting non-trivial variables are discussed, including: mean difference, eta squared, cumulative probability, and half-normal plots. How to handle subjects in an experiment is described when their characteristics can be included as experimental factors and when they cannot be, and when subjects are merely a form of replication. The values of introducing multiple center points into the 2^{k-p} fractional factorial screening designs are discussed, particularly as they are used to build the screening design into a central-composite design. Lack-of-fit tests are provided to help decide whether a second- or a third-order response surface is needed. Numerous methods of analyzing screening studies with multiple responses are described. A method is given for developing a prediction equation with data collected from an incompleted screening design.

ACKNOWLEDGMENTS

This report was prepared under contract with the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under Prime Contract No. F44620-76-0008 with Canyon Research Group, Inc. Contract monitor was Dr. Alfred Fregly.

As usual, the work could not have been accomplished without the help of a great many people. Major contributors are acknowledged here:

Elizabeth Lage Roscoe was Research Assistant during the early phases of the program. She contributed materially in the exploration of ways to combine advanced methodologies into optimized plans, in preparing tabular material and numerical examples, and in collecting and organizing references.

Howard B. Lee not only made the contributions specifically noted in the text but also provided important references that otherwise might have been overlooked and helped clarify some of the more obtuse statistical papers.

Marilyn Wilson Lewis provided significant help in the preparation of the final tables and examples used to illustrate the techniques.

Rose E. Konrath and Masse Bloomfield of the Hughes Aircraft Company Library made the total effort easier by the assistance they gave collecting original papers and reports for the project.

Paula Lintz provided many valuable comments regarding the editing of the final report.

Elaine Benveniste contributed to the final preparation of the report.

Others who made special contributions specifically noted in the text are: Edward J. Dragavon, Seymour Singer, and Steve B. Webb.

ACKNOWLEDGMENTS

This report was prepared under contract with the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under Prime Contract No. F44620-76-0008 with Canyon Research Group, Inc. Contract monitor was Dr. Alfred Fregly.

As usual, the work could not have been accomplished without the help of a great many people. Manor contributors are acknowledged here:

Elizabeth Lage Roscoe was Research Assistant during the early phases of the program. She contributed materially in the exploration of ways to combine advanced methodologies into optimized plans, in preparing tabular material and numerical examples, and in collecting and organizing references.

Howard B. Lee not only made the contributions specifically noted in the text but also provided important references that otherwise might have been overlooked and helped clarify some of the more obtuse statistical papers.

Marilyn Wilson Lewis provided significant help in the preparation of the final tables and examples used to illustrate the techniques.

Rose E. Konrath and Masse Bloomfield of the Hughes Aircraft Company Library made the total effort easier by the assistance they gave collecting original papers and reports for the project.

Paula Lintz provided many valuable comments regarding the editing of the final report.

Elaine Benveniste contributed to the final preparation of the report.

Others who made special contributions specifically noted in the text are: Edward J. Dragavon, Seymour Singer, and Steve B. Webb.

FOREWORD

This report is one of a series written by me under a common title: "Advanced methodologies for human factors engineering research." It covers one segment of the total data collection and analysis process in a complete research program, namely, that phase dealing with the screening of a very large number of variables to discover the critical ones. According to the research strategy that I am trying to promote, the screening process is not a complete experiment and should only be used after a thorough analysis of the real world is made to develop a list of candidate variables to be screened and before a later effort to develop a complete and accurate response surface. Screening designs and response surface designs are not two separate designs, but the first is a first stage of the second; the second is an outgrowth of the first.

This report describes a screening process that is an improvement over that written in the earlier reports, integrating economical multifactor research techniques with those that keep the data relatively free from trend effects. Use of this report presumes that the reader is already familiar with the earlier reports, particularly those on economical multifactor designs, on building trend-robust designs, and on ridge regression analysis, as well as the basic principles for conducting economical behavioral research. New ideas for improved analysis and for handling multiple response data are introduced here.

The techniques discussed in this report are treated unevenly. Forced by time limitations to either go into considerable detail regarding a small piece of the screening process, or provide an overview of the complete process, I chose the latter. Even for the more sparsely treated techniques, I have tried to present enough information that would not only direct the reader's attention to potentially useful methodologies but, by judicious sifting and digesting, would also help clarify the original papers when they are read. Only one important step -- data transformation -- was omitted because I was not satisfied that the method I had would do the job properly.

Eventually, the missing details will have to be added, along with more details on the other phases of the research process after screening. Although experience is needed to determine the full power of this approach, merely studying 20, 30, or 40 variables in a systematically-manipulated experiment cannot help but improve the predictive quality of the research or the generalizability of the data base.

Charles W. Simon
1977

TABLE OF CONTENTS

Section Page

I. INTRODUCTION 1

EXPERIMENTAL DESIGN 4

II. 2^{p-q} RESOLUTION IV SCREENING DESIGN PLANS 5

CHARACTERISTICS OF THE SCREENING DESIGNS IN THIS REPORT 6

Multifactor 6
Economical 6
Quasi-saturated. 6
Two-level factors. 7
Resolution IV. 7
Trend-robust 7
Factor-level-change sensitive. 7
Robust to experimenter error 8

CONSTRUCTING RESOLUTION IV SCREENING DESIGNS . . 8

From Resolution III Designs 8

Plackett and Burman designs 11

Complete Resolution IV Designs 12

Resolution IV Designs Robust to Trend 13

Rearranging the columns 14
Identifying the experimental conditions. . 18
Identifying the trend-robust columns . . . 18
Counting factor-level changes 24

PREPARING TO USE SCREENING DESIGNS 26

Pre-analysis To Select the Experimental Factors 26

Pre-analysis To Facilitate the Use of Screening Designs 29

Pre-experiment Data Collection 30

Selecting the Screening Design 33

Resolution III or IV designs 35
Number of factors 36

Table of Contents (Continued)

Page

The Designs 36

Assigning Factors to the Columns of the Design 37

Trends 38
Count 39

Modifying the Given Designs 41

When Fewer than the maximum possible number of factors are studied 44

III. EFFECTIVE USE OF CENTER POINTS IN SCREENING DESIGNS 46

CENTRAL-COMPOSITE DESIGNS 48

SCALING 49

QUALITATIVE FACTORS 51

IV. INTRODUCING SUBJECTS INTO THE SCREENING DESIGN AS FACTORS AND AS REPLICATIONS 53

SUBJECT CHARACTERISTICS AS EXPERIMENTAL FACTORS 53

Measuring Subject Characteristics as Part of the Design 53

Measuring Subject Characteristics Not in the Design 55

SUBJECTS AS REPLICATION 57

Replicating for Inter-subject Reliability . . 58

Replicating to Establish Confidence Intervals 59

DATA ANALYSIS 60

V. CALCULATING CRITERIA TO SELECT NON-TRIVIAL FACTORS 61

SELECTION CRITERIA 61

Estimating the Effects 63

Table of Contents (Continued)

Page

Yates algorithm 66
Estimating the coefficients of the multiple regression equation 71
Interpreting effects data 72

Estimating the Proportion of Variance Accounted For 73

Eta squared 74
Interpreting proportion of variance data 74

Cumulative Proportion of Variance 75

Reverse Yates' algorithm 78
Interpreting the cumulative effects of non-critical factors 80

Half-normal Plots (Daniel) 83

Preparing half-normal grids 85
Plotting the data 89
Interpreting half-normal plots 89
Interpretation tactics 90
Detecting defective values 92

Standardized Half-normal Plot (Zahn) 93

USING ORDERED DISTANCES WITH MULTIPLE RESPONSE DATA 99

VALIDATION TEST 100

NUMERICAL EXAMPLE OF A SCREENING STUDY ANALYSIS 102

VI. ANALYZING SUBJECTS AS REPLICATIONS 109

ESTIMATING CONFIDENCE LIMITS 110

INTERPRETING MULTI-SUBJECT DATA 111

VII. ADJUSTING EXPERIMENTAL EFFECTS FOR TRENDS 115

CONSTRUCTING AN ARTIFICIAL PROBLEM WITH TREND EFFECTS 116

DETERMINING THE VALUE OF TREND-ADJUSTMENT FACTORS 121

Linear Adjustment Factor $\dot{L}$ 121

Table of Contents (Continued)

Page

Quadratic Adjustment Factor $\dot{Q}$ 124

Cubic Adjustment Factor $\dot{K}$ 124

Linear Plus Cubic Adjustment Factor 125

Making the Adjustment for Trend 126

Applying Trend-adjustment Techniques to Screening Designs 128

VIII. HANDLING MULTIPLE RESPONSES IN SCREENING STUDIES . 129

ADVANTAGES OF BILATERAL MULTIVARIATE EXPERIMENTS 129

Scope of This Section 132

WEIGHTED CRITERIA 133

GRAPHIC INSPECTION 134

USING LAGRANGE MULTIPLIERS 136

STEP-DOWN PROCEDURE 138

MULTIPLE ANALYSIS OF VARIANCE (MANOVA) 140

MANOVA Eta Squared 142

One-way designs 142
Multifactor (and screening) designs . . . 144

Multi-variate Test of Significance 147

MANOVA Versus Multiple Discriminant Analysis 149

GRAPHICAL ANALYSIS USING ORDERED DISTANCES . . . 151

General Description 152

Gamma distribution paper 153

Calculating the Ordered Distances 154

Selecting the compounding matrix 155

Table of Contents (Continued)

Page

Analyzing Subgroups 157

Plotting and Evaluating the Ordered Distances 157

Conclusion 157

CANONICAL CORRELATION ANALYSIS 158

Applications 158

Process 159

Limitations of Canonical Correlation Analysis 162

An Improved Method of Canonical Correlation Analysis 163

IX. EVALUATING THE ADEQUACY OF THE REGRESSION EQUATION 166

SPLITTING THE LACK OF FIT OF THE FIRST ORDER DESIGNS 167

SPLITTING THE LACK OF FIT OF SECOND ORDER DESIGNS 169

Calculations 170

EFFECTS OF REPLICATING NON-CENTER POINTS OF THE CCD 172

ADDITIONAL CRITERION FOR EVALUATING THE EQUATION 172

Calculation and Test 173

X. ANALYZING THE DATA FROM AN INCOMPLETE SCREENING EXPERIMENT 176

REQUIREMENTS FOR USING THE P-C EQUATION 177

PREDICTOR-CORRECTION EQUATION 177

EXAMPLE 178

MISSING DATA 184

XI. REFERENCES 185

APPENDICES 191

LIST OF TABLES

Table No. Page

1 2^{8-4}_{IV} TREND RESISTANT SCREENING DESIGN WITH FACTOR-LEVEL CHANGE COUNT AND PERCENT OVERLAP . . 15

2 ILLUSTRATING USE OF TCHEBYCHEFF'S COEFFIENENTS TO CALCULATE INNER-PRODUCT SUMS AND SUM OF SQUARES 21

3 EFFECTS ON TREND-EFFECT OVERLAP AND FACTOR-LEVEL CHANGE COUNT WHEN DESIGNS ARE MADE WITH MAIN EFFECTS ASSOCIATED WITH A, B, C, OR D FACTORS . . 43

4 CENTER POINTS IN AN EXPERIMENTAL DESIGN INVOLVING QUANTITATIVE AND QUALITATIVE FACTORS 52

5 CONVENTIONAL METHOD OF ANALYZING 2^k DESIGN 65

6 SIGN MATRIX METHOD OF ANALYZING 2^k DESIGN 65

7 YATES' ALGORITHM FOR ANALYZING A 2^k FACTORIAL . . 67

8 USING A SPLIT-YATES' ALGORITHM TO ANALYZE A 2^k DESIGN 70

9 PROBABILITY VALUES FOR CONSTRUCTING HALF-NORMAL GRIDS 87

10 THE ELEVEN FACTORS AND THE TWO LEVELS ASSIGNED EACH FACTOR 103

11 EXPERIMENTAL CONDITIONS AND PERFORMANCE SCORES FOR NAVAL WEAPONS CENTER STUDY 104

12 ANALYSES OF NAVAL WEAPONS CENTER EXPERIMENTAL DATA 105

13 METHOD OF CONSTRUCTING ARTIFICIAL TREND-BIASED DATA 117

14 ESTIMATING EXPERIMENTAL EFFECTS IN THE ABSENCE AND PRESENCE OF TREND EFFECTS 120

15 EQUATIONS NEEDED TO FIND THE LINEAR, QUADRATIC, AND LINEAR/CUBIC CORRECTION VALUES 122

16 SYMBOLS USED IN EQUATIONS TO CALCULATE L'_2 SUM OF SQUARES FOR THE SECOND-ORDER CENTRAL-COMPOSITE DESIGN 171

17 IMAGINARY DATA WITH WHICH TO ILLUSTRATE AN INCOMPLETE ANALYSIS 179

18 WORKING DATA TO OBTAIN UPDATED EQUATIONS 179

LIST OF FIGURES

Figure No.		Page
1	Examples of Trial-to Trial Performance Variability	31
2	Examples of Types of Interactions Including No Interaction	34
3	Illustrating How Single Center Point Enables Each Factor To Be Tested At Three Levels . . .	47
4	Plotting Brightness on a Linear and Logarithmic Scale	50
5	Illustration of How Experimental Context (i.e., Task Difficulty) Affects Performance . .	72
6	Half-normal Grids for Plotting Thirty-one Absolute Effects	88
7	Half-normal Plots on Results of Naval Weapons Center Experiment	107
8	Comparing Results from Imaginary Experiment (Table 13), Both Trend-free and Trend-biased .	118
9	Artificial Data Illustrating Graphic Overlapping of Two Response Surfaces	135
10	Illustration Showing How Analysis of Single Responses in Multiple-response Experiments May Fail to Detect Real Differences	141
11	Geometric Representation of One Main Effect, A, in a 2^5 Experiment with Two Responses . . .	154

I. INTRODUCTION

Screening designs -- a class of fractional factorials -- are systematic data collection plans that enable the effects of a very large number of factors to be estimated economically. Screening designs are used primarily in the second phase of a total research program where they are intended to determine which of the great many factors have non-trivial effects on the performance of a particular task. Screening designs are to be used to identify important factors, not to obtain an accurate representation of the experimental space. This latter operation will occur in subsequent phases of the research program.

The strategy for using screening designs in this manner stems from the observation that a great many psychological and human factors experiments investigate trivial factors. Simon (1975b), in an analysis of 239 experiments published in *Human Factors* over a fourteen year period, found that in experiments studying from one to five factors, 24 percent of the 494 main effects examined accounted for *one percent* or less of the total variance in the experiment. Forty-one percent of main effects accounted for only four percent of the total variance in the experiment.

As might be expected, the more factors included in a single experiment, the more frequently trivial effects were found. Similar conditions have been found in analyses of other journals that publish psychology experiments (Gallo, *et al*, 1977; Dunnette, 1966, p 35).

With the great many factors that are likely to affect performance in any given task, one must wonder why any psychologist, interested in predicting and controlling performance, would study factors having trivial effects.

Why not first study the factors accounting for the large effects? The principle of maldistribution (Budne, 1959; Simon, 1973; 1976b) leads us to expect that a relatively few factors account for most of the variance. These should be investigated first in order to build a structure of data within which marginal effects can be located and about which confidence limits can be established.

Of course, the answer to "why?" is that until the experiments are completed, one would not know which factors are important. But this is where screening designs become applicable. Instead of doing many three- or four-factor experiments, with highly replicated designs, requiring a great many observations to collect redundant information of limited value, the screening designs provide a means of examining a great number of factors with the maximum amount of information with a minimum amount of redundance and relatively few observations. What the results from many little traditional experiments cannot do, but which results from the screening design can, is to order the factors according to the size of their effects and to discover interactions among factors that appear within the same experiment. Screening designs do all this economically for they can be used to study N factors with 2N observations (although the size of the designs in this report will all be equal to some power of 2). Thus if there are 25 factors, for example, to be ranked in terms of their importance, only 64 observations would ordinarily be required when screening designs are used. Furthermore, the precision with which the main effects are estimated is usually much greater than the effects measured in many smaller, yet highly replicated studies. The effects obtained from screening studies not only permit the ranking of factor effects on a quantitative scale, but can provide an equation approximating the experimental space if that space can be represented by a linear model.

The beauty of using a screening design is that once the important factors have been identified (step one), the same data can be used, if supplemented by relatively few additional observations at new experimental conditions, to complete a response surface (step two) capable of accurately approximating the experimental space defined by the original set of 25 factors. For several hundred observations, a reasonable approximation of a 25-factor space is possible. These capabilities arise through the appropriate application of the principles of economical multifactor research (Simon, 1973), the basic strategy being to collect only the data needed to supply the information required at each particular phase of the research program. Screening designs are employed in the second phase to help (in as economical an effort as possible) the investigator decide what factors, what measures, what range of values should be investigated in greater detail at a later stage of the program.

18

EXPERIMENTAL DESIGN

- How to design Resolution IV screening designs robust to linear, quadratic, and cubic trend effects without replicating the basic design.

  Complete designs are provided requiring 8, 16, or 32 observations to quantitatively order the effects of up to 8, 16, or 32 factors.

- How to prepare to use screening designs: preliminary empirical studies and analyses.

- How to assign operational factors to the design to keep them robust to trend while minimizing the number of difficult or time-consuming level changes.

- How to add center points to a screening design to roughly estimate error variance and to provide the data needed to test how well a linear model fits the empirical data.

- How to include multiple subjects in the screening design: dimensionalized as factors.

II. 2^{p-q} RESOLUTION IV SCREENING DESIGN PLANS

In this section, how to construct a special type of screening design and the preparations recommended for using them will be described. The section is written with the assumption that the reader is familiar with the information on fractional factorials in general and screening designs in particular, as described by Simon (1973) in an earlier report, or its equivalent. The reader should also be familiar with certain techniques for constructing trend-free 2^k designs, which may be found in Simon (1974) or the original papers. The techniques described in those two reports are consolidated in this report to provide an extremely economical and efficient experimental design for identifying critical factors.

Although the methods of construction are described here, three complete screening designs are provided in this report in spite of a strong personal belief by the author that "cookbook" applications of experimental plans are to be deplored and are bound to degrade the quality of research in the long run. Cookbook applications enable the uninformed to mimic the efforts of qualified investigators enough, in many cases, to provide a face validity to their efforts while masking sloppy data collection, an inadequate analysis, and a misinterpretation of results. They allow the lazy investigator to fit his problems to his methods and his experiments to the designs that are available in a book, rather than to design each experiment in a way that is likely to provide the most valid information needed for the problem at hand.

The justification for providing these ready-made designs, therefore, lies mainly in their utility in illustrating the design principles described in this report and in reducing the amount of routine calculations an investigator would have to perform in developing the designs on his own. Proper use of the designs still requires a great deal of involvement by the investigator in order to fit them to his problem.

CHARACTERISTICS OF THE SCREENING DESIGNS IN THIS REPORT

Each design exhibits the following characteristics:

1. Multifactor. A single run of these designs can be used to estimate the effects of up to 8, 16, or 32 factors. By analogy, still larger designs can be constructed. However, in practice, if adjustments for trend effects are to be made, one degree of freedom for each order of trend (i.e., linear, quadratic, or cubic) must be set aside, reducing the number of experimental factors that can be studied.

2. Economical. The effects of up to K factors can be estimated with N observations, when K equals N/2 and N equals some power of 2 (e.g., 2^4, 2^5, 2^6). The designs in this report require 16, 32, and 64 experimental conditions in a single run for studying up to 8, 16, and 32 factors, respectively.

3. Quasi-saturated. The designs allow for no independent estimate of the error term unless one wishes to assume that two-factor interaction strings are negligible. If fewer than the maximum possible number of factors are studied, the effects of three-factor interaction strings can be estimated. Without additional information, it would be incautious to assume them to be equivalent to an independent estimate of error.

4. Two-level factors. These designs sample only two levels of a factor, although they could be adapted to handle four levels per factor if necessary (Cochran and Cox, 1957, p 273). However, since these plans are to be used for screening, about the only justification for a four-level factor would be when there are four conditions of a qualitative variable. The two levels would be selected near psychophysical or practical performance limits of the factor to measure the full effect.

5. Resolution IV. All main effects can be isolated from one another and from all two-factor interactions. Each main effect will be aliased with a different string of three-factor interactions. Two-factor interactions will be aliased with one another in isolated strings.

6. Trend-robust. The experimental conditions of each design are ordered so that without replication, estimates of many main effects will be totally unaffected by linear, quadratic, and cubic trends - for example, subject learning or equipment drift - confounded with the effects of interest. All but a few effects will be robust to trends. The designs are arranged so that it is easy to identify the more trend-robust columns to which factors are assigned.

7. Factor-level-change sensitive. If the levels of a factor are difficult or time-consuming to change, the investigator may use the change-counts provided with each design to assign the difficult-to-change factor to a column requiring few changes.

8. Robust to experimenter error. These 2^{k-p} designs are remarkably robust to variations in setting the experimental conditions of the independent variables, even when the experimenter is unaware of the existence of the error (Box, 1963).

9. Modular. Center-points and additional levels for each factor can be added to the designs to provide the data needed to estimate non-linear, quadratic effects of a second-order response surface. New blocks of experimental conditions can be added to the original Resolution IV design to create Resolution V designs that form the center of a central-composite design.

CONSTRUCTING RESOLUTION IV SCREENING DESIGNS

Since screening designs are merely a form of the 2^{k-p} fractional factorial designs, they can be constructed in a number of different ways. Several methods in addition to the one used for the plans in this report are described in order to provide the user with the greatest degree of flexibility of method.

From Resolution III Designs

Simon (1973, pp 89-116) explains the techniques developed by Box and Hunter (1961) and Daniel (1962) for constructing Resolution IV screening designs from two Resolution III designs. A Resolution III design is constructed by first writing down the sign matrix for the full factorial and then aliasing additional factors with the interactions of the original design. For example, a seven-factor Resolution III design with eight observations would be constructed by aliasing new factors with the interactions of a 2^3 factorial plan, thus:

	Column Headings							
Original 2^3 factorial:	(I)	A	B	C	AB	AC	BC	ABC
2^{7-4}_{III} created by aliasing:	(I)	A	B	C	D	E	F	G

With this design, N-1 main effects can be isolated from one another but not from two-factor or higher interactions. The defining generators are:

$$(I) = H = ABD = ACE = BCF = ABCG$$

The research strategy would be to collect and analyze the data from the conditions of this first block (a Resolution III design) in order to discover if the design, the factors, and the range of conditions are adequate and to make whatever changes are needed before collecting additional data. When a great many factors are being investigated, information from this single block may be sufficient in some cases to drop some of the variables before commencing data collection on the second block.

When the investigator is ready to collect more data, he constructs a second design composed of experimental conditions for a second Resolution III block that are the "foldovers" of the first block. In the foldover design, the levels of all conditions -- including (I) = Factor H -- are reversed. The defining generators for this second block would be:

$$(I) = -H = -ABD = -ACE = -BCF = ABCG$$

The defining generators for the combined design can be derived by expanding each set of generators into the full set of defining contrasts and adding the two sets together. Box and Hunter (1961, p 338) provide a rule that simplifies the process. They write:

> *. . . when a design is formed containing 2^{k+1} runs from a design containing 2^k runs by replicating the 2^k design with reversed signs and associating some further factor X with the 2^k plus ones and 2^k minus ones, then a general rule for obtaining the generators and defining relation of the new design from the generators and defining relations of the old design is as follows: 1) All generators which contain an even number of characters in the original design are retained as generators in the new design, 2) All generators which contain an odd number of characters in the original designs will be reproduced containing the extra character X as generators in the new design.*

Thus, in our example, when the two Resolution III designs are combined, the result is a Resolution IV design with the following defining generators:

$$(\mathrm{I}) = \mathrm{ABDH} = \mathrm{ACEH} = \mathrm{BCFH} = \mathrm{ABCG}.$$

The defining contrasts (or defining relations as Box and Hunter call them) are obtained by expanding the defining generators by multiplying all combinations of the original generators in pairs, triplets, and so forth. For the above example, the complete set of defining contrasts would be:

$$(\mathrm{I}) = \overset{1}{\underline{\mathrm{ABDH}}} = \overset{2}{\underline{\mathrm{ACEH}}} = \overset{3}{\underline{\mathrm{BCFH}}} = \overset{4}{\underline{\mathrm{ABCG}}} = \overset{12}{\mathrm{BCDE}} = \overset{13}{\mathrm{ACDF}} = \overset{14}{\mathrm{CDGH}} =$$

$$\overset{23}{\mathrm{ABEF}} = \overset{24}{\mathrm{BEGH}} = \overset{34}{\mathrm{AFGH}} = \overset{123}{\mathrm{DEFH}} = \overset{124}{\mathrm{ADEG}} = \overset{134}{\mathrm{BDFG}} = \overset{234}{\mathrm{CEFG}} = \overset{1234}{\mathrm{ABCDEFGH}}$$

where the numbers above each contrast indicate which of the defining generators (underlined) it is a product. Since the resolution of the design can be determined by the number of letters in the smallest defining contrast, it is apparent that the two Resolution III designs, when combined, form a Resolution IV plan.

Plackett and Burman designs. Resolution IV designs also can be made from the Plackett and Burman (1946; also see Simon, 1973, pp 102-104) Resolution III designs by adding an additional "foldover" block. One advantage of using those designs would be the extra economy achieved as the number of factors to be studied increases. This economy derives from the fact that the Plackett and Burman designs can be constructed by restricting the number of experimental conditions to some multiple of four. The Box and Hunter designs, on the other hand, require that the number of experimental conditions be restricted to some power of two. Thus, if one wished a Resolution IV design for fifty factors, the Box and Hunter designs would require two Resolution III blocks of 64 (or 128) experimental conditions while Plackett and Burman designs would require two blocks of 52 (or 104) experimental conditions. Another advantage of Plackett and Burman designs for screening purposes was noted by Tukey (1960, p 171), who found that the degree of confounding between main and two-factor interaction effects in the Resolution III Plackett and Burman plans was quite low in many cases (and much less than the fully aliased conditions in the Box and Hunter designs). Estimating the relative strength of main effects with the Plackett-Burman designs before continuing to the foldover block might, therefore, be done with greater confidence. Neither the Plackett-Burman designs nor their potential applications will be discussed further in this report. The reader, however, should consider using them if they fit his problem.

Complete Resolution IV Designs

The designs proposed in this report do not provide for a progressive data-collection plan in which a Resolution III design is used first to investigate the linear effects (aliased with all higher order effects) to be followed by a second block to isolate main and two-factor interaction effects. Instead, with these designs, it is presumed that the isolation of main and two-factor interaction effects is an absolute requirement for screening purposes and so all the data for that purpose is collected at one time.

Box and Hunter (1961, p 341) note that a Resolution IV design can be constructed directly "by first writing down the sign matrix for a two-level factorial and then associating new variables with all interaction columns having an odd number of [letters]." Thus, a 16-observation Resolution IV design can be derived from a 2^4 factorial plan by aliasing four new factor labels (e.g., E,F,G, and H) to the four three-factor interactions (i.e., ABC, ABD, ACD, and BCD) in the original plan. By the proper assignment of new factor labels, this design can be made equivalent to the design made from the principal fraction plus foldover Resolution III designs described in the previous section.

The reader should be aware by this time of a number of characteristics common to all of these methods. The sign matrix for any design formed from a factorial plan is arranged so that row coefficients are orthogonal among themselves, as are column coefficients among themselves.* With rows representing the independent experimental conditions, sources of variance can be assigned to the columns in various combinations. A column may be labeled a main effect or an interaction, or as with saturated designs, a string of interactions. However, whatever label is assigned to a column, since columns are orthogonal, we may be certain that an effect measured in any one column *will be* independent of an effect measured in any other column. Thus, we may label the columns as we please, as long as we are careful to see that labels for the main effects and those for their interactions are assigned consistently with the requirements of the sign matrix. With these principles in mind, a screening design robust to trend can be created.

---

*With the plus and minus signs actually representing plus and minus ones, orthogonality between any pair of columns can be checked by obtaining the cross-product sum between columns, which must equal zero. The same is true with rows.

Resolution IV Designs Robust to Trend

Two steps are required to construct the designs provided in this report. The first is to construct a quasi-saturated fractional factorial that will be suitable for screening purposes. The second is to adapt it so as to take advantage of its trend-resistant characteristics.

We begin to construct the design by first determining the design size which depends on the number of factors being investigated. The rule is:

> The number of experimental conditions required is the nearest power of two (2^k) that is equal to or greater than twice the number of factors to be studied.

For example, we wish to study 20 factors. Two times twenty equals 40. The nearest 2^k equal to or greater than 40 is $2^6 = 64$ conditions. Or, perhaps we wish to study 8 factors. Eight times two equals 16. The nearest 2^k value is $2^4 = 16$ conditions.

For this example we shall construct a screening design to study eight factors. First it is necessary to lay out the sign matrix for a complete 2^k factorial design. For this example we use a sign matrix for a 2^4 factorial design. There would be 16 (N) experimental conditions, arranged in the Yates' (1937) "standard order," capable of estimating the following (N-1=15) effects, also arranged here in the standard order:

A,B,AB,C,AC,BC,ABC,D,AD,BD,ABD,CD,ACD,BCD,ABCD

plus the mean (I). These are referred to in this paper as the "old" or the "original factorial" labels.

Rearranging the columns. We rearrange the column of signs by moving all columns with labels that include Factor A* to the left and all remaining columns to the right. The effects with Factor A are then ordered from the largest to the smallest interactions followed by the main effect, A. Also, within any order of interaction, they would be arranged alphabetically. For example, this would be:

$$\begin{array}{ccccccccl}
\multicolumn{4}{c}{\overbrace{\qquad\qquad\qquad\qquad\qquad\qquad}^{\text{Alphabetical}}} & \multicolumn{3}{c}{\overbrace{\qquad\qquad\qquad}^{\text{Alphabetical}}} & & \\
\underline{\text{ABCD}}; & \underline{\text{ABC}}, & \underline{\text{ABD}}, & \underline{\text{ACD}}; & \underline{\text{AB}}, & \underline{\text{AC}}, & \underline{\text{AD}}; & \underline{\text{A}} & \text{(New labels)} \\
4 & 3 & 3 & 3 & 2 & 2 & 2 & 1 & \text{(Size of effect)}
\end{array}$$

The reason for this particular arrangement will be more evident later. These steps can be followed from here on by examining the completed design in Table 1.

---

*Selecting Factor A for this purpose is arbitrary. Later, in order to find columns that are robust to trends and also require few factor level changes, it may be necessary to use a different factor.

TABLE 1

2^{8-4}_{IV} TREND RESISTANT SCREENING DESIGN WITH FACTOR-LEVEL CHANGE COUNT AND PERCENT OVERLAP

TEST ORDER	EXPERIMENTAL CONDITION	NEW SCREENING DESIGN LABELS*															
		(MAIN EFFECTS)*									(TWO-FACTOR INTERACTION STRINGS)**						
		(I)	A	B	C	D	E	F	G	H	AH	AG	AF	AE	AD	AC	AB
1	AEFG	+	+	-	-	-	+	+	+	-	-	+	+	+	-	-	-
2	BCDH	+	-	+	+	+	-	-	-	+	-	+	+	+	-	-	-
3	BCFG	+	-	+	+	-	-	+	+	-	+	-	-	+	+	-	-
4	ADEH	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
5	BDEG	+	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-
6	ACFH	+	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-
7	ACDG	+	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-
8	BEFH	+	-	+	-	-	+	+	-	+	-	+	-	-	+	+	-
9	CDEF	+	-	-	+	+	+	+	-	-	+	+	-	-	-	-	+
10	ABGH	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+
11	ABDF	+	+	+	-	+	-	+	-	-	-	-	+	-	+	-	+
12	CEGH	+	-	-	+	-	+	-	+	+	-	-	+	-	+	-	+
13	ABCE	+	+	+	+	-	+	-	-	-	-	-	-	+	-	+	+
14	DFGH	+	-	-	-	+	-	+	+	+	-	-	-	+	-	+	+
15	(I)	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
16	ABCDEFG	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ORIGINAL FACTORIAL LABELS		(I)	ABCD	ABC	ABD	ACD	AB	AC	AD	A	BCD	BC	BD	CD	B	C	D
PERCENT # TREND/EFFECT OVERLAP ***	LINEAR									1					5	19	75
	QUADRATIC						0	1	5			5	18	71			
	CUBIC			1	3	10				3	41				10	21	11
FACTOR-LEVEL CHANGE COUNT		0	10	11	9	13	8	12	14	15	5	4	6	2	7	3	1

*THREE-FACTOR INTERACTION STRINGS ALIASED WITH MAIN EFFECTS ARE LISTED IN APPENDIX I-A.

**TWO-FACTOR INTERACTION STRINGS ALIASED WITH TWO-FACTOR INTERACTION LABELS LISTED IN APPENDIX I-B.

***INNER-PRODUCT SUMS LISTED IN APPENDIX I-C.

#BLANK SPACES REPRESENT ZERO PERCENT. SPACES WITH ZEROES REPRESENT SOME PERCENT SMALLER THAN 1%.

Next, we assign "new screening" labels, i.e., the letters from A to H (for the eight factors in our design), to the rearranged columns which still bear the old factorial labels, thus:

New labels (Screening Design):	A	B	C	D	E	F	G	H
Original labels (Factorial Design):	$\underline{ABCD}$,	$\underline{ABC}$,	$\underline{ABD}$,	$\underline{ACD}$,	$\underline{AB}$,	$\underline{AC}$,	$\underline{AD}$,	$\underline{A}$

These are not aliases in the usual sense; instead they are merely associations that occur from the relabeling. To minimize confusion, all original factorial labels, hereafter, will be underlined.

We must next arrange the columns in which Factor A is not present in the original factorial labels. This is done by first arranging the columns from left to right according to the order of the old labels (from the highest to the lowest interaction and then the main effects), and within each order, arrange the effects alphabetically. In our example, the columns would be arranged like this:

$$\begin{array}{lllll l}
\underline{BCD}, & \overbrace{\underline{BC},\ \underline{BD},\ \underline{CD}}^{\text{Alphabetical}}, & \underline{B}, & \underline{C}, & \underline{D}, & \text{(Old label)} \\
3 & 2 \quad\ \ 2 \quad\ \ 2 & 1 & 1 & 1 & \text{(Number of factors involved)}
\end{array}$$

There is one less term than there was in the previous set with the Factor $\underline{A}$. The missing column is the Identity column, (I).

Next we must associate new screening labels with these old ones. All new ones will be two-factor interactions of the A to H new labels given to the other set. It happens that when columns are arranged so that their original factorial labels are as shown above, new label two-factor interactions including A will be arranged in reverse alphabetical order thus:

AH, AG, AF, AE, AD, AC, AB

This makes the column of the original label BCD the column of the new label AH; the column with the original label BC is now the column with the new label AG, and so forth. The complete association across all 16 columns then would be:

New:	(I)	A	B	C	D	E	F	G	H	AH	AG	AF	AE	AD	AC	AB
Old:	(I),	ABCD,	ABC,	ABD,	ACD,	AB,	AC,	AD,	A,	BCD,	BC,	BD,	CD,	B,	C,	D

To show that the column associated with both AH and BCD (new and old labels) is the appropriate one for the interaction between the columns associated with A or ABCD and H or A, we multiply new and old at the same time. The associations remain consistent, thus:

	New		Old
	A	=	ABCD
Multiplied by	H	=	A
Yields	AH	=	BCD

This would be true with any of the other combinations. With the new labels, the 2^4 factorial design has been turned into a 2^{8-4} screening design, since all main effects, being in different columns, are orthogonal to themselves and to all two-factor interactions.

The next step is to find the aliases within the strings of two-factor interactions. The simplest procedure is to continue the pairing of factors, this time beginning with B, i.e., BH, BG, BF, BE, BD, BC, and not repeating any previously used pair, e.g., BA=AB. This makes the number of pairs get smaller each time around, i.e., AH to AB, BH to BC, CH to CD, DH to DE, EH to EF, FH to FG, and GH. There will be k(k-1)/2 combinations for K factors. For the fully quasi-saturated design, each string of two-factor interactions will contain k/2 interactions. For example, AH would be aliased with (in this example) DE (since ACD x AB = BCD); CF (since ABD x AC = BCD); and BG (since ABC x AD = BCD). Aliases are provided for the designs given in this report. A computer program for identifying aliases, prepared by Mr. Howard Lee, is given in Appendix IV.

Identifying the experimental conditions. The columns, along with their old and new labels, have been rearranged. For the old labels, the names of the experimental conditions remain the same. For the new labels, new names of the experimental conditions must be obtained. This can be done with the newly arranged sign matrix. Each row is a different (and independent) experimental condition. The "name" of each experimental condition can be obtained by writing down a letter corresponding to each new label main effect in the rearranged design that has a plus sign under it in the particular row. It is conventional to write the names of experimental conditions in small letters leaving capital letters for the names or labels of the effects of the columns. For example, if the first row of the sign matrix looked like this after the rearrangement:

New labels:	(I)	A	B	C	D	E	F	G	H	AH	AG	etc
Signs :	+	+	-	-	-	+	+	+	-	-	+	etc

then the experimental condition associated with that row would be:

aefg

since the letters correspond to those of the main effects with + signs in their columns.

Identifying the trend-robust columns. The reasons for the particular column arrangement described above will now become more evident. The general idea on which this is based came from a paper by Daniel and Wilcoxin (1966, p 261; also see Simon, 1973, pp 121-128) who noted that:

> *. . . certain of the* *ordered* *contrasts appearing in the 2^p system are orthogonal to linear and to quadratic trends. Some other contrasts are nearly orthogonal and some are rather heavily correlated with first and second order trend. The design problem is, then, to choose those sets of ordered contrasts that provide efficient estimation of all desired effects and interactions.*

What they are saying is that certain columns, (i.e, the vertical sequences of plus or minus coefficients, in a sign matrix of a two-level factorial or fractional factorial experimental design) arranged with the experimental conditions in standard order, correlate zero or very little with a set of coefficients representing a linear or a quadratic trend. The same is true for cubic trends, which Daniel and Wilcoxin did not consider in their paper. The investigator would want to assign the more important factors to the column most robust to trend so that estimated effects would not be distorted.

Other methods (see Simon, 1973) for handling sequence effects have been proposed. Some involve making multiple measures of each condition and arranging them in sequences that eventually are balanced against trends. Some methods require a large number of repeated measures in which the effects have been introduced randomly and the trend effects isolated by means of statistical techniques. Both approaches involve far more data collection than is usually justified during the early screening process. The method proposed by Daniel and Wilcoxin (1966) provides the most economical solution by taking advantage of the natural robustness to trend of 2^{k-p} or 2^k designs, unreplicated.

To determine the degree to which each column of our screening design is robust to linear, quadratic, and cubic trend effects, we must correlate the plus and minus (one) coefficients in each column of the sign matrix with the appropriate integer Tchebycheff orthogonal polynomial coefficients (Fisher and Yates, 1963; Beyer, 1966; DeLury, 1950).

Let us illustrate this with the column for Factor G in the 2^{8-4} screening design (Table 1), originally labeled Interaction AD in the factorial plan. The ordered column vector of coefficients (without the ones) for Factor G, and the ordered Tchebycheff coefficients for linear, quadratic, and cubic trends are shown in Table 2. The correlation (r) between linear (L) trends and Factor G is obtained thus:

$$r_{LG} = \sqrt{\frac{(\Sigma LG)^2}{(\Sigma L^2)\ (\Sigma GG)}}$$

where ΣLG^2 is the sum of the cross products (or inter-product sum) between each pair of effect and linear trend coefficients

ΣLL is the sum of the linear trend coefficients, each squared

ΣGG is the sum of the squared coefficients for Factor G (which will equal N in these designs)

Thus to calculate the values needed to solve the equation, from the data in Table 2, we do the following:

$$\Sigma LG = (-15)(+1)+(-13)(-1)+(-11)(+1)+\ldots(+13)(-1)+(+15)(+1)=0$$
$$\Sigma GG = (+1)^2 = (-1)^2 + (+1)^2 \ldots (-1)^2 + (+1)^2 = 16$$
$$\Sigma LL = (-15)^2 + (-13)^2 = (-11)^2 \ldots (+13)^2 + (+15)^2 = 1,360$$

TABLE 2

ILLUSTRATING USE OF TCHEBYCHEFF'S COEFFICIENTS TO CALCULATE INNER-PRODUCT SUMS AND SUM OF SQUARES

	Factor G Coefficient*	TCHEBYCHEFF'S COEFFICIENTS		
		Linear	Quadratic	Cubic
	+	-15	+35	-455
	-	-13	+21	- 91
	+	-11	+ 9	+143
	-	- 9	- 1	+267
	+	- 7	- 9	+301
	-	- 5	-15	+265
	+	- 3	-19	+179
	-	- 1	-21	+ 63
	-	+ 1	-21	- 63
	+	+ 3	-19	-179
	-	+ 5	-15	-265
	+	+ 7	- 9	-301
	-	+ 9	- 1	-267
	+	+11	+ 9	-143
	-	+13	+21	+ 91
	+	+15	+35	+455
ΣGG	16	1360	5712	1007760
ΣLG		0		
ΣQG			64	
ΣKG				0

---

*Plus or minus signs represent coefficients of +1 and -1 respectively.

Substituting these values in the equation, we get:

$$r_{LG} = \sqrt{\frac{0^2}{16 \times 1260}} = \sqrt{\frac{0}{21760}} = 0$$

With a zero correlation, an estimated effect of Factor G would be totally unaffected if an unwanted linear trend effect was running through the data.

Repeating the process for the quadratic trend and Factor G we get:

$$\Sigma QG=(+35)(+1)+(+21)(-1)+(+9)(+1)+\ldots(+21)(-1)+(+35)(+1)=64$$
$$\Sigma QQ=(+35)^2+(+21)^2+(+9)^2+\ldots(+21)^2+(+35)^2=5712$$
$$\Sigma GG=N=16$$

Substituting in the equation, we get:

$$r_{QG} = \sqrt{\frac{64^2}{16 \times 5712}} = \sqrt{\frac{4096}{91392}} = .044818 = .2117$$

The percentage of overlap between the quadratic trend and the effect of Factor G is, therefore:

$$\%_{QG} = (r_{QG})^2 \times 100 = (.2117)^2 \times 100 = .0448 \times 100 = 4.5$$

The correlation between Factor G and the cubic trend effect was zero.

To discover which columns are the most robust to trends, this process is repeated for all relationships between linear, quadratic, and cubic trend effects and the experimental effects (main and two-factor interaction strings). However, these calculations are supplied for the designs given in this report.

Daniel and Wilcoxin (1966, pp 269-270) point out how Yates' (1937) algorithm, when applied directly to the Tchebycheff coefficients, can be used to calculate the innerproduct sums more easily than if these were obtained a column at a time.

When all of the effects for any design are correlated, the relationships show two distinct patterns. For one, referring to the original factorial labels, certain types of sources are always correlated with particular trend effects. Thus:

Four-factor interactions and higher	Uncorrelated with L, Q, or K trends*
Three-factor interactions	Correlated with cubic but not with linear or quadratic
Two-factor interactions	Correlated only with quadratic
Main effects	Correlated with linear and cubic but not quadratic

A second pattern is also apparent. Within any set of effects of the same order, if they exist at all, the correlations increase (using the labels of the original factorial) as the factors progress alphabetically. Thus, the AC interaction would be more correlated with a quadratic effect than the AB interaction, and so forth. Both of these patterns can be seen in the 2^{8-4}_{IV} design (Table 1), but they become even clearer with larger designs.

It should be clearer now why the columns of the screening design are reordered as they are. It allows main effects (new labels) to be assigned to the columns less correlated with trends and the two-factor interaction strings to be assigned to columns more correlated with trend effects. For screening purposes, this greater emphasis on keeping main effects clean is appropriate. The column reordering also tends to place the least correlated within these two groups more to the left of the design, facilitating its use.

---

*The letter K is used to represent the cubic trend to avoid confusion when the letter C is used to represent an experimental factor.

Since this general pattern is not completely correct, with each of the designs given in this report the percentage overlap (= r x 100) between each factor and trend combination is provided. The investigator can use these when he must decide how to assign real-world factors to the design columns.

Counting factor-level changes. One can merely count the number of times any column requires a change of factor levels. For example, in the 2_{IV}^{8-4} design (Table 1), in column AB the level is changed only once, from low to high between the eighth and ninth trial, while in column H, the levels are changed fifteen times, every other trial. Within each design, the number of times the factor level changes (the count) is a different value in each column, from one to N-1 for N experimental conditions (and N-1 effects) in each 2^{k-p} design.

As the designs get bigger, it may be inconvenient, as well as time-consuming to count the changes in each column. The following algorithm can be used instead:

1. Using the original factorial labels, with experimental conditions in Standard Order, determine the counts for the main effects. If the letters for the main effects are written in reverse alphabetic order, the count for each will be:

   $$(2^k - 1)$$

   where k is the position of the main effect in the reverse order sequence:

   e.g., in a three-factor design, the main effects are A, B, and C. In reverse order they are *C*, *B*, and *A*, in positions 1, 2, and 3 respectively. Their factor level change count would therefore be:

$$\begin{array}{ll} \underline{C}: & 2^1 - 1 = 1 \\ \underline{B}: & 2^2 - 1 = 3 \\ \underline{A}: & 2^3 - 1 = 7 \end{array}$$

2. To determine the count for any interaction, the counts for the individual main effects are combined always as: plus, minus, plus, minus, etc, starting with plus and going as far as necessary;

   e.g., the count for the interaction *ABC* -- the letters must be ordered alphabetically -- would be:

$$\begin{array}{cccccccc} & \underline{A} & & \underline{B} & & \underline{C} & & \\ + & 7 & - & 3 & + & 1 & = & 5 \end{array}$$

   or for *BC*:

$$\begin{array}{cccccc} & \underline{B} & & \underline{C} & & \\ + & 3 & - & 1 & = & 2 \end{array}$$

   or for *AB*:

$$\begin{array}{cccccc} & \underline{A} & & \underline{B} & & \\ + & 7 & - & 3 & = & 4 \end{array}$$

   of course, the count can be simplified since *ABC* would also equal:

$$\begin{array}{cccccc} & \underline{AB} & + & \underline{C} & & \\ + & 4 & + & 1 & = & 5 \end{array}$$

PREPARING TO USE SCREENING DESIGNS

It takes more to properly design an experiment than to describe the experimental design. Screening designs tell us at what coordinates in the abstract experimental space we should sample performance to obtain information regarding main effects without bias from two-factor interaction effects. However, the investigator has more to do if he wishes to use these designs effectively.

Pre-analysis to Select the Experimental Factors

Before he selects the final set of factors to be included in the screening study, the investigator should prepare an unrestricted list of factors which reasonable and knowledgeable experts believe may have a non-trivial influence on the real-world task most of the time. This first step is designed to make certain that any source likely to influence the performance of the task under investigation be listed for consideration, whether it be related to the equipment, subject, environment, or task. The value of this exercise is to reduce omissions too early in the effort because of practical considerations, real or imagined, at that time. This, of course, is no license to list every factor imaginable, but any that are likely to influence the performance at hand should be included in this initial step.

The second step is to define the task, with emphasis on the conditions in the real world. This includes an operational definition of the performance measure (and more likely, measures) that will be employed, as well as the nature of the stimuli and responses of the specific situation. While this short statement does not do justice to the care required and the importance of this requirement, the matter will not be discussed any further in this report.

The third step is to decide what real-world values to set at the upper (+1) and lower (-1) limits of each factor. These values should be selected, based on the following considerations:

1. Limits likely to be experienced in the real world for the task under consideration.

2. Limits set by the state-of-the-art in the real world.

3. Limits set by the state-of-the-art in simulation, which may be beyond those in the real world so that information regarding future systems can be collected.

4. Limits set by construction costs, where the information lost is not considered critical.

5. Limits set by manipulation difficulties, where the information lost is not considered critical.

6. Limits that are likely to approximate the points at which the highs and lows of performance will occur. (This is particularly important when the function between the factor and performance is probably U-shaped.)

The limits selected obviously affect, to some extent, how critically a particular factor will appear to affect performance. If the limits are too narrow, performance may change little and an investigator may read this (incorrectly) as meaning the factor has a trivial effect on performance, when in fact it is true only within the limits being studied. Had the limits been set wider, the effect would be greater. This is why setting the limit values should be determined by real-world interests, so that effects are measured under conditions of practical interest in the operational situation. Do not do as one eminent psychologist did when he failed to get an effect from some factor. That is, do not expand the range in the simulation beyond anything likely to be found in reality so that the factor would show a significant effect.

The fourth step is to assign priorities to the original list of variables based on a number of considerations:

1. Order the factors on a five-point scale (if possible) according to how much each -- within its specified limits -- is likely to affect the performance on the particular task.

2. Indicate those factors in which the interested parties (e.g., contracting agency and the investigator) have a special interest.

3. Indicate those factors that are expensive to simulate.

4. Indicate those factors that are likely to interact with one another, noting particularly the ones likely to result in disordinal interactions.

The investigator must weigh these subjectively to select the final set of the factors for the experiment. The listing exercise provides him with a better overview when making his decision. Ultimately, he must consider how his decision affects the experiment's capacity to reflect reality for the task under consideration.

Pre-analysis to Facilitate the Use of Screening Designs

Once the factors have been selected, the next step is to anticipate how they will fit into a screening study. This can be a mixed process of analysis and empirical data collection. However, pre-analysis is always desirable, whether or not it is to be followed by preliminary or formal data collection, for it can show *a priori* that certain effects, observed later, were anticipated. An anticipated disordinal interaction can be accepted as real with greater confidence when found in the data than one that was not anticipated.

The investigator should make the following analyses as an aid to using the screening design:

1. *Classify the factors according to their quantitative characteristics: ordered-continuous; ordered-discrete; ordered-complex-categorical (by choice); categorical*.

   This provides a preview of design characteristics needed to handle each factor. Ordered factors can eventually become part of a response surface, and may be assumed continuous for certain applications, but all levels of the factor may not be available as a design data collection point. On the other hand, complex factors which are treated as categorical ones but which are in fact a particular combination of ordered and continuous factors, may be redefined according to these parameters. Most economical multifactor designs can be used more effectively with ordered and continuous factors; fewer data points generally have to be taken and the chances that greater-than-second-order interactions are non-trivial are small.

For the ordered factors,

2. Estimate the response function between the given limits of each factor. Four functions are of major interest: linear, quadratic, U-or negatively accelerated growth pattern, cubic or S-shaped. This will aid in deciding how complex a model may be needed to approximate the response surface, how many levels will be needed to approximate the individual functions, and where the limiting data points must be located.

3. Decide what measurement scale might be used to simplify any non-linear function that was anticipated. This helps meet the requirements for a lower-order response surface when economical multifactor designs are used.

4. Attempt to draw the interaction effects that are considered important, and consider the scaling that would eliminate the ordinal interactions.

Pre-experiment Data Collection

Certain information can only be obtained empirically. Some data might be collected, if deemed important by the investigator, to make a quick but tentative check on assumptions made in the foregoing analysis. Other information, however, is vital if the screening designs are to be used, and should be collected, in fact, prior to any experiment. The most important ones are:

1. Test the trial-to-trial reliability for a single condition. (Reliability Test)

Test a typical subject on five or more consecutive trials of a single condition. Does performance from trial to trial vary irregularly and excessively (see Figure 1-A)? If so, this suggests that some critical source of variance has not been identified and/or is not under control, and should be. Are there signs of a progressive trend effect over the five trials (see Figure 1-B)? This suggests that the subject might not be sufficiently familiarized with the task or the experimental apparatus. Either more practice, trend isolation techniques, or both may have to be employed. Is there an immediate improvement in performance and then a leveling off (see Figure 1-C)? This suggests that some precautions need be taken to offset momentary perturbations as each new trial condition is introduced.

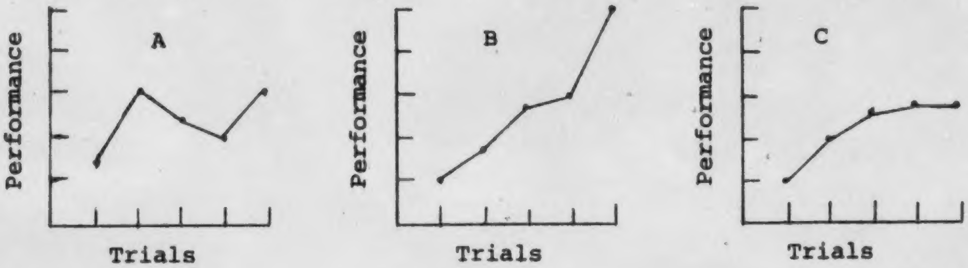


Figure 1. Examples of Trial-to-Trial Performance Variability

Differences in the extent of trial-to-trial variance under easy and difficult conditions with experienced and inexperienced subjects provide clues to the need for proper response scaling and other variance-control mechanisms.

2. Test for subject-to-subject variability within presumably homogeneous groups. (Subject Heterogeneity Test)

On both easy and difficult experimental conditions, a number of presumed equivalent subjects should be tested. If their performance differs considerably, then one may suspect that critical subject characteristics are being ignored. Quite often, subjects are considered homogeneous according to some simple label, but are not so insofar as their performance is concerned. This test provides some clues as to whether those subject factors should be measured or controlled in the experiment.

When faced with the need to introduce a new subject characteristic as a dimension of a screening design, the investigator must consider the nature of the characteristic. If the characteristic is simple and readily quantifiable (e.g., visual acuity), then it probably should be introduced into the experimental design as any other factor. This means that subjects within different levels of visual acuity -- two levels for a screening study -- would be used, each performing a particular combination of the levels of the remaining factors representing the experimental condition. If the characteristic is complex and difficult to quantify (e.g., pilot experience), initially it might be better to run subjects representing each of the two levels on every experimental condition. This permits a subject-by-factor interaction, if it exists, to be detected.

3. Test to determine whether there are conditions which can be performed perfectly or can't be performed at all by most subjects.

When too many experimental conditions are too difficult or too easy, the information provided by a screening design is severely limited. An investigator may have to "live with it," or he may find that by making slight adjustments in the range of a few factors, he can eliminate these uninformative upper and lower limits. This, however, should never take priority over practical interests and the reality of the situation.

4. Test a very good and a very poor subject on the easiest and most difficult tasks. (Interaction Test).

How performance is distributed among these four conditions provides valuable clues regarding the task, its range of difficulties, and the scaling of the dependent variable. Four types of solutions are shown in Figure 2. In Figure 2-A no interaction is present, while in Figure 2-B an important type of disordinal interaction is shown, warning of the presence of interactions that the unaugmented screening design is poorly equipped to handle. Figures 2-C and 2-D suggest the presence of ceiling and floor effects, respectively, which may be reduced through appropriate scaling.

Selecting the Screening Design

Two major considerations in selecting a screening design are:

1) whether or not one wishes to isolate main and two-factor interaction effects immediately before examining part of the data;
2) the number of factors to be studied.

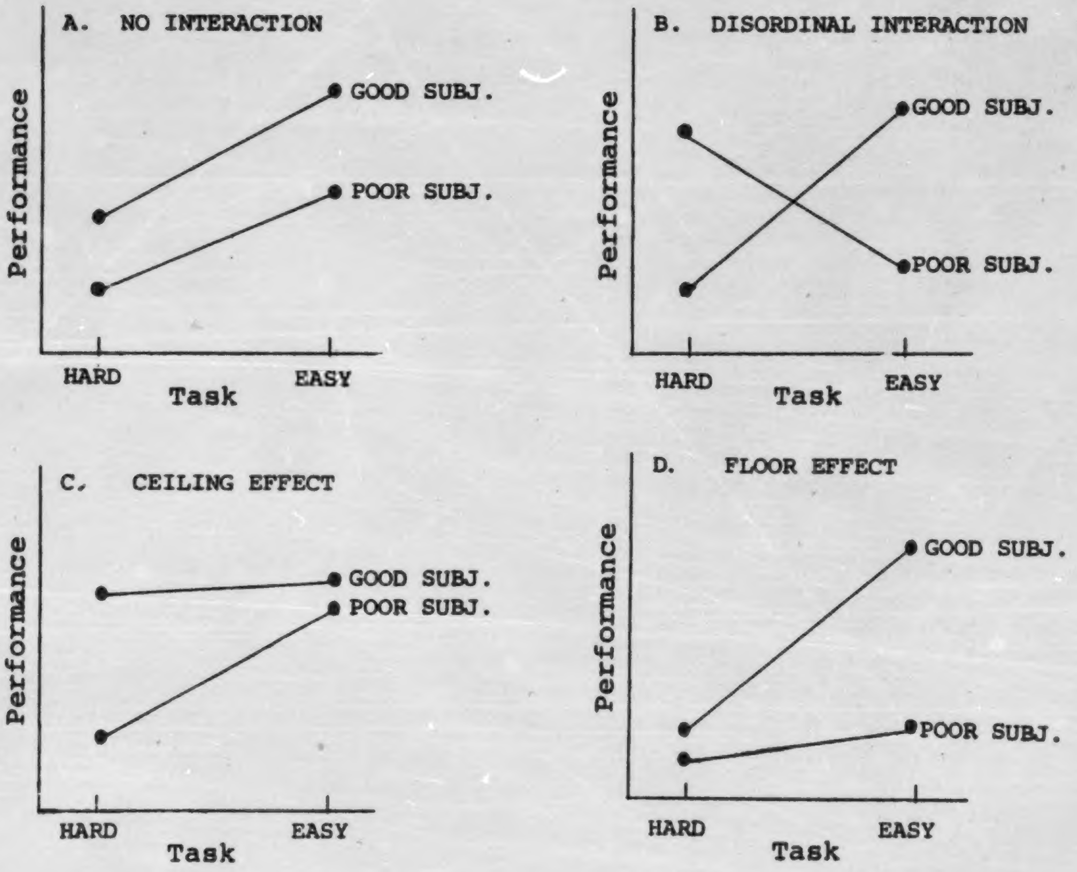


Figure 2

Examples of Types of Interactions Including No Interaction

Resolution III or IV designs. It would be unusual for an investigator performing human factors engineering experiments (or any behavioral science study) not to want to isolate main from two-factor interaction effects. Two-factor interactions occur too frequently to risk their distorting main effects, even in a screening study. On that basis, an investigator may wish to use a Resolution IV design from the beginning without resorting to blocking. The added advantage of the Resolution IV design is robustness to trends (more so than a Resolution III design).

As more factors are to be investigated and the cost of data collection becomes uncomfortably high, there may be stronger reasons to begin with a Resolution III design, as the first block, and then later add a second Resolution III design to create a Resolution IV design. First of all, blocking enables data to be examined and factors added or dropped, or their ranges changed if necessary, after half as much data has been collected as would be required were the full Resolution IV design completed first. Second, blocking facilitates the control of certain irrelevant sources of variance (Simon, 1970a; 1974 pp 100-103). Finally, running experiments in small blocks reduces the chances that some disruptive force would destroy the entire experiment. Equipment breakdowns may be less likely to occur and subject sickness may be easier to avoid. In either case, it is easier to recoup from the loss of a small block of data than it would be if an entire study were lost because of some disturbance occurring part way through the experiment.

In this report, only the Resolution IV plans will be discussed. Resolution III plus foldover plans were discussed in an earlier report (Simon, 1973, pp 89-125).

Number of factors. The Resolution IV designs provided in this report are capable of handling up to 8, 16, or 32 factors (and others capable of handling up to 2^6, 2^7 ... 2^n factors can be created by the same process). However, the number of experimental factors that can be studied in any design will be reduced if the investigator plans to isolate trend effects or is restricted by particular combinations of factors, interactions, and trend contamination.

The investigator must allow for making trend estimates, losing one column (or experimental factor) for each trend (linear, quadratic, and/or cubic) effect that is to be isolated. Further restrictions on the number of available columns (and therefore factors to be studied) may occur if an investigator wants to keep certain combinations of main and interaction effects robust to linear or quadratic effects. If he decides to block his design, he must sacrifice still more columns, which reduces the number of factors that can be studied still more. For all of these reasons, an investigator must select a design large enough to handle more than just the number of experimental factors.

The Designs

The following basic Resolution IV screening designs and supporting data are provided in this report:

Design	Number of Factors to be Studied	Minimum Number of Observations (N) for a Single Replication	Design to Use
2^{8-4}_{IV}	Up to eight	16	Table 1, Appendix I
2^{16-11}_{IV}	Nine to 16	32	Appendix II
2^{32-26}_{IV}	Seventeen to 31	64	Appendix III

The following information is provided with each design:

1. The sign matrix
2. The experimental conditions
3. The original factorial design labels
4. The new screening design labels
5. Trend-robust test order
6. Percentage overlap between linear, quadratic, and cubic trend effects and experimental design effects
7. Number of changes made between levels for each factor
8. Two-factor interaction aliases
9. Three-factor aliases of main effects
10. Inter-product sums used to adjust factors for trend effects

Assigning Factors to the Columns of the Design

In assigning the real-world factors to the columns of the design matrix, the investigator will be concerned with which main effects and which interactions must be kept trend-robust and which will require the fewest number of factor-level changes. These decisions, of course, will depend on:

a) Which ones are the most important and thus should be estimated with the smallest amount of trend bias.

b) Which ones are important but are so unquestionably large that they will be identified even though the data is somewhat distorted.

c) Which ones are the most difficult cr most time-consuming to change from level to level.

d) Which ones are likely to show large two-factor, disordinal interaction effects.

Trends. In each table, the percentage overlap at the bottom of the columns shows the investigator how much each column will be contaminated with trend effects. Columns affected by linear trends are not affected by quadratic trends. In making his selection, however, the investigator should realize that in human factors performance data, linear effects are generally larger than quadratic, and both are generally larger than cubic effects. Thus, a 10% overlap for a linear effect would ordinarily be much more likely to distort the data than a 10% quadratic overlap. There are, of course, no absolute rules and the investigator is obligated to minimize these effects by his experimental procedures (Simon, 1974, pp 21-26) so that when trends do appear, relative to the effects under investigation, they will be small to begin with, making the absolute amount of overlap even smaller.

Special problems of assignment arise when the investigator wishes to keep both main effects and the two-factor interactions reasonably trend-free. There are fewer interaction columns that are trend-free or trend-robust, and the magnitude of the overlap is, on the average, higher than in the main effect columns. Anything not overlapping more than 10% with a linear trend is probably reasonable to use.

An overlap of less than 30% and 50% between interactions and quadratic or cubic effects, respectively, would also probably be acceptable if the investigator had no reason to believe that this type of trend would be present to any degree and had done his best to reduce them through his data collection procedures. These percentages are of course arbitrary, and depend in part on how cautious an investigator feels he must be.

As the number of factors increases, i.e., the larger the designs, the options available to an investigator in this regard, increase. Even if the investigator can't get a trend-free interaction column with these designs, he still has two options. First, he can make adjustments for trends (to be discussed in the section on Analysis). Second, he may modify the design (to be discussed later in this section).

Count. Screening designs are valuable because they permit a large number of factors to be investigated quickly. But if it takes a great deal of time to change the factor levels from trial to trial, this prime advantage will be lost. The sophisticated experimenter -- if he has any say in the matter -- will see that every means is taken when the experimental apparati are being built to insure that a rapid and accurate change can be made between levels of all factors. Delays may affect the subject's motivation and performance, and errors in settings can destroy the value of the data. When normal precautions are taken, however, it is more common to find that only a few of the total number of factors have serious difficulties insofar as changing the factor levels is concerned.

The problem of assigning the factors to the proper columns of the design depends, therefore, on both the number of factors that must be considered as well as the degree of difficulty in making the factor-level changes. For example, if it takes a day to make a change in the level of a particular factor, then the investigator would probably prefer to assign the main effect of that factor to a column requiring only a single change. If it takes only several minutes, he may be content to assign it to a column requiring more changes.

Unfortunately, with the designs provided in this report, the main effects are all associated with columns that require at least N/2 or more changes, where N is the number of experimental conditions in the design. This means that even the main effect column with the smallest factor-level change count still requires a great many changes. Furthermore, this problem increases as the size of the design increases.

The problem of factor assignment is further complicated if the investigator wishes the column selected for its minimum number of changes also to be reasonably robust to trend effects. But it is apparent from the designs, that, on the average, those columns most robust to trends are the ones requiring the greatest number of factor-level changes. The designs, as they have been arranged for this paper, maximize this inverse relationship. For example, in the 2^{16-11}_{IV} design (Appendix II) the column identified as the string containing the AB interaction is the one requiring only a single factor-level change, but it also is the one with the cubic trend. The column requiring only two factor-level changes (i.e., the string of two-factor interactions with AF in it) has a 71% overlap with the quadratic trend, a somewhat better situation, but not a comfortable one. About the first reasonable compromise in the 32-run design would be the column identified as the two-factor interaction string including AH, requiring four factor-level changes and an overlap of only 4% with the quadratic trend.

Thus, it seems that with the designs given in this report, in order to have only a few factor-level changes, a main effect must be assigned to one of the columns made up of two-factor interaction strings. While this is possible, since it has already been noted that we may assign any labels to the columns, it is still not a simple matter, for it triggers a series of reactions involving the other columns in order to maintain the appropriate relationships among main and interaction effects. However, there is a solution that an investigator may use if necessary. The given designs are intended to optimize the robustness to trends, but if it is also necessary to be concerned with factor-level changes at the same time, the designs can be easily modified to meet this need.

Modifying the Given Designs

With the given designs, the smallest factor-level change count for a main effect will be equal to N/2, where N is the number of observations in the study, and in the 2^{8-4}_{IV} designs, no linear nor quadratic trend effect overlaps a main effect by 10% or more. If it is necessary to reduce the factor-level count, by sacrificing the robustness to trend, one may repeat the procedures given in this report to create the original designs in Table 1 and Appendices II and III except that instead of assigning to main effects all of the columns containing a Factor A in the original labels, we would assign all those containing Factor B, or Factor C, or Factor D and so forth in the original labels, instead, depending on what mixture of factor-level count and trend resistance is required.

For example, with the 2^{16-11}_{IV} design, if we used all columns originally labeled with Factor B in them for the main effects, then the smallest factor-level change count associated with a main effect would be four, but now only one out of eight main effects is overlapped by linear or quadratic trends by more than 10%. If all columns labeled with Factor *C* in them had been used for the main effects, then the smallest count would be two and only two of the eight factors would overlap linear and quadratic trend effects by more than 10%. At the same time that trend-resistance among main effects is decreasing, more trend-resistant columns are being associated with the two-factor interaction strings. The effects of building a 2^{16-11}_{IV} screening design where the main effects are associated with the columns of the original Factors *A*, *B*, *C*, or *D* are shown in Table 3*.

---

*For completeness, the reader should be aware of other efforts to develop experimental plans that are robust to trend while minimizing the number of factor-level changes required. Simon (1974, pp 138-146) described the methods proposed by Draper and Stoneman (1968) and Dickinson (1974). Their plans were limited in two ways: 1) they were robust only against linear time trends and 2) their robustness was only for main effects. They arrived at what they believed were optimum designs through a systematic examination of each alternative; this becomes increasingly expensive as the size of the design increases and it also reduces experimenter options. Joiner and Campbell (1976) proposed to reduce the costs by searching optimum combinations of a random subset of the various alternatives. Lancaster and Reynolds (1976) proposed a method whereby the investigator could select the optimum combinations for both main and interaction effects.

TABLE 3

EFFECTS ON TREND-EFFECT OVERLAP AND FACTOR-LEVEL CHANGE COUNT WHEN DESIGNS ARE MADE WITH MAIN EFFECTS ASSOCIATED WITH A, B, C, OR D FACTORS

A

New Screening Design Labels		A	B	C	D	E	F	G	H	1	2	3	4	5	6	7
Original Factorial Labels		ABCD	ABC	ABD	ACD	AB	AC	AD	A	BCD	BC	BD	CD	B	C	D
% Trend/ Effect Overlap	Lin.								1					5	19	75
	Quad.					0	1	5			5	18	71			
	Cub.		1	3	10				3	41				10	21	11
Factor-level Change Count		10	11	9	13	8	12	14	15	5	4	6	2	7	3	1

B

New Screening Design Labels		A	B	C	D	E	F	G	H	1	2	3	4	5	6	7
Original Factorial Labels		ABCD	ABC	ABD	BCD	AB	BC	BD	B	ACD	CD	AD	AC	A	C	D
% Trend/ Effect Overlap	Lin.								5					1	19	75
	Quad.					0	5	18			71	5	1			
	Cub.		1	3	41				10	10				3	21	11
Factor-level Change Count		10	11	9	5	8	4	6	7	13	2	14	12	15	3	1

C

New Screening Design Labels		A	B	C	D	E	F	G	H	1	2	3	4	5	6	7
Original Factorial Labels		ABCD	ABC	ACD	BCD	AC	BC	CD	C	ABD	AB	AD	BD	A	B	D
% Trend/ Effect Overlap	Lin.								19					1	5	75
	Quad.					1	5	71			0	5	18			
	Cub.		1	10	41				21	3				3	10	11
Factor-level Change Count		10	11	13	5	12	4	2	3	9	8	14	6	15	7	1

D

New Screening Design Labels		A	B	C	D	E	F	G	H	1	2	3	4	5	6	7
Original Factorial Labels		ABCD	ABD	ACD	BCD	AD	BD	CD	D	ABC	AB	AC	BC	A	B	C
% Trend/ Effect Overlap	Lin.								75					1	5	19
	Quad.					5	18	71			0	1	5			
	Cub.		3	10	41				11	1				3	10	21
Factor-level Change Count		10	9	13	5	14	6	2	1	11	8	12	4	15	7	3

Once the columns have been rearranged in an order that produces a satisfactory factor-level change count and trend resistant combination, it is necessary to assign the new screening design labels. If the design is arranged in the same manner described when Factor A terms were used, then all new screening-design labels for both main and interaction effects and their aliases will remain the same.

Finally, the new experimental conditions must be renamed because when the columns have been reordered and assigned to different main effects, the order in which the experimental conditions will occur will also change. This is accomplished by merely writing down the letters (using small letters for the conditions) associated with all main effects with a plus sign in each row.

When fewer than the maximum possible number of factors are studied. The designs in Table 1 and Appendices I, II and III are suitable for investigating up to a maximum of 8, 16, and 32 experimental factors, respectively, less of course the number set aside to handle trend or blocking. Quite often, however, an investigator will not want to investigate the maximum number possible, and will want to modify the given designs accordingly. This is done by simply striking out each letter representing the label of each unused column from the letter designations of the experimental conditions, and by removing all interactions in the strings of aliases containing those letters. This may create an uneven number of interactions among the strings.

For example, in Table 1, if there were only six factors in the experiment and no G and H factors were used, then the experimental conditions would be changed as follows:

With G & H	Without G & H
aefg	aef
bcdh	bcd
bcfg	bcf
adeh	ade
bdeg	bde

etc.

and Main Effect A would only be aliased with BCE, and BDF and Interaction AF would only be aliased with BD, since all interactions with the letters G and H in them would be eliminated.

The columns in which no main effects are located are now used to estimate directly the effects of particular strings of three-factor interactions.

III. EFFECTIVE USE OF CENTER POINTS IN SCREENING DESIGNS

Unreplicated 2^{k-p} screening designs have two distinct limitations: 1) they cannot measure possible curvilinear relations between independent and dependent quantitative variables; 2) they provide no direct estimate of the experimental error variance. These are recognized, but to obtain such information would be costly and, for screening purposes, would be of little value and certainly not cost-effective. In later stages of research, this information does become important. Since the investigator can ordinarily anticipate continuing his experiment beyond the screening phase, he would fit a non-linear model, if necessary, and obtain an external error estimate at that time.

Data from at least three levels of each continuous factor is needed to measure the curvature of the response surface. Screening designs ordinarily have only two levels. Design points must be replicated several times to estimate error variance. Replication for this purpose is usually discouraged in the screening study. However, once the decision is made to get this information, it can be obtained most economically by adding data-collection points at the center of the experimental design. Data collected at the center of the design (with coded coordinates 0,0..., 0, when the original screening design coded coordinates were +1 and -1) will provide some estimate of curvilinearity for every factor.

By adding a single point at the center of the experimental design, a third -- middle -- level of every factor of the screening design is measured. This is illustrated in Figure 3. With three levels, -1, 0, and +1, performance at

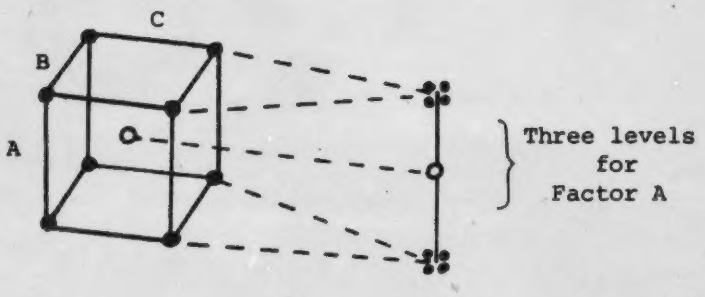


Figure 3

Illustrating How Single Center Point Enables Each Factor to be Tested at Three Levels.

each end point is estimated by averaging one half of the experimental conditions in the original screening design.

The center position, however, would be estimated from the performance of only the single center point. Because of this uneven precision along the dimension, with the poorest being at the center of interest, repeated measures should be made at the center of the design. This center-point replication will also provide an empirically derived estimate of the experimental error.

CENTRAL-COMPOSITE DESIGNS

Box and Hunter (1958) propose using this center-point replication technique in their central-composite designs, where there are still more advantages than indicated above. Since central-composite designs follow in the research program once the critical factors* have been screened, multiple center points should be included in the screening designs whenever appropriate. The number of center-points in central-composite designs affect the following design characteristics and functions:

1. The test for the presence of quadratic effects in the first-order model and higher-order effects in the second-order model.

2. The estimate of "pure" error variance needed to test the statistical significance of lack of fit.

3. The uniformity of the "information" profile (which is based on the number of observations at each point in the response surface).

4. The orthogonality of the central-composite design.

---

*An optimum design strategy would use the data collected in the screening design as a block of data making up the cube portion of the central-composite design. The methodology for handling this transition will not be discussed in this report.

5. The "rotatability" of the central-composite design.

6. The ability to isolate block and trend effects.

As applied to the central-composite design, the above items are discussed in considerable detail by Box and Hunter (1958, pp 152-168) and Simon (1974, p 102; 1976a, pp 22-28). Lack-of-fit tests can be applied to screening designs supplemented with multiple center points. These will be discussed in the Analysis section of this report.

SCALING

Once an experimenter has decided to add center points to the screening design, he is forced for the first time to consider what measurement scale to use for each factor. Up to now, since basic screening designs are made up with two levels per factor, only a linear response surface could be estimated regardless of what shape might actually exist in the real world. Adding center points complicates the situation.

Let us, for example, consider a 2^3 factorial study involving Sensor Resolution (5 and 15 feet), Target Brightness (10 and 100 foot-lamberts), and Vehicular Speed (300 and 600 knots). The pairs of values set the limits of the three-dimensional experimental space. The experimenter who decides to add center points should not automatically select the point with coordinates in the center of each dimension, i.e., 10 feet resolution, 55 foot-lamberts brightness, and 450 knots vehicular speed. Instead, he should first consider what scale will enable the experimental space to be represented by as simple a function as possible.

To illustrate this, let us consider the brightness scale. With a center point at 55 foot-lamberts, experience has shown that the scale would relate non-linearly with a visual performance task (Figure 4A). On the other hand, when brightness data is plotted on a logarithmic scale, the relation would more nearly approximate a straight line (Figure 4B).

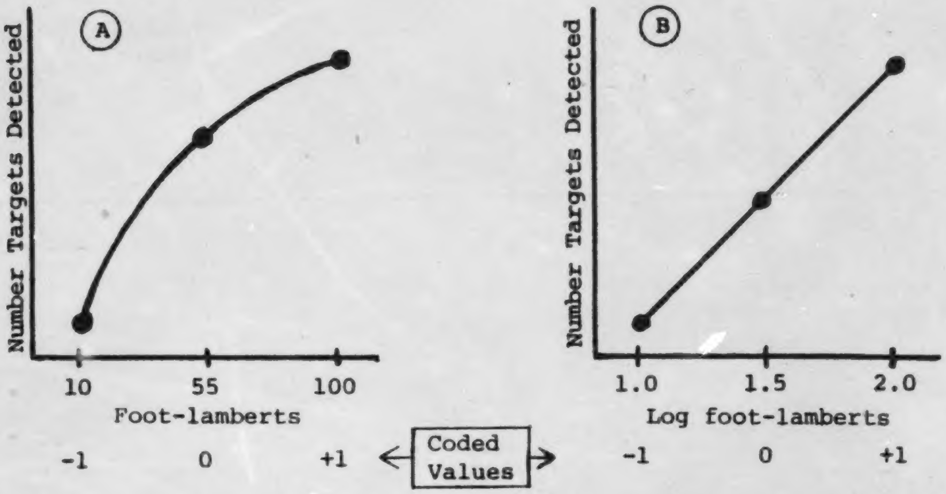


Figure 4.
Plotting Brightness on Linear and Logarithmic Scales

Since economical multifactor research is most successful the simpler the relationship and since fewer conditions need be studied to approximate the less complex functions, the experimenter would be better off using a log foot-lambert scale while maintaining the range between 10 and 100 foot-lamberts (i.e., one and two log foot-lamberts). This means that the center point on that scale would be at 1.5 log foot-lamberts, or 31.6 foot-lamberts instead of 55 foot-lamberts.

A similar decision must be made for the Vehicular Speed. The experimenter would want to consider whether speed or rate (the reciprocal of one another) is likely to give the simplest function. The choice will determine whether 450 knots or its reciprocal in seconds would be used.

QUALITATIVE FACTORS

Center points can be added to a design only when the factors are quantitative and continuous. Categorical variables have no order and therefore no center values. However, when quantitative and qualitative variables are studied in the same experiment, center points can still be added. In that case, the condition would be centered only within the space defined by the quantitative, continuous variables. This restricted center point would be replicated once for each unique combination of the qualitative variables.

This is illustrated in Table 4. A sign matrix is given for an experiment with two qualitative and two quantitative factors. The first sixteen conditions are those of a full 2^4 factorial, with + and - representing the coded +1 and -1, high and low values. The last four conditions show how center points (0,0 in the coded terms) for the quantitative variables are added.

67

TABLE 4

CENTER POINTS IN AN EXPERIMENTAL DESIGN INVOLVING QUANTITATIVE AND QUALITATIVE FACTORS

	FACTORS				
	Qualitative		Quantitative		
Experimental Conditions*	I	II	III	IV	
1	-	-	-	-	Original factorial design
2	+	-	-	-	
3	-	+	-	-	
4	+	+	-	-	
5	-	-	+	-	
6	+	-	+	-	
7	-	+	+	-	
8	+	+	+	-	
9	-	-	-	+	
10	+	-	-	+	
11	-	+	-	+	
12	+	+	-	+	
13	-	-	+	+	
14	+	-	+	+	
15	-	+	+	+	
16	+	+	+	+	
17	-	-	0	0	Minimum number of additional center points
18	+	-	0	0	
19	-	+	0	0	
20	+	+	0	0	

---

*This is not intended to be an optimized presentation order.

IV. INTRODUCING SUBJECTS INTO THE SCREENING DESIGN AS FACTORS AND AS REPLICATIONS

Subjects in psychological experiments either appear as 1) identifiable types who can be represented as composite levels of subject factors, or as 2) unidentified masses, presumed to be homogeneous members of the same population.

SUBJECT CHARACTERISTICS AS EXPERIMENTAL FACTORS

There are two situations that can exist when we wish to include subject characteristics as factors along with equipment/environment factors and temporal factors. In one, each subject is selected having the characteristics required by the sign matrix. In the other, measurable subject characteristics are known but it is difficult to impossible to select subjects with the required combinations.

Measuring Subject Characteristics as Part of the Design

If each subject characteristic were to be investigated at two levels, and there are f characteristics, 2^f subjects would be required to exhibit all of the required combinations of characteristics. Each subject would be tested on a particular combination of the remaining factor levels, where the combined characteristics of subjects and other factors would represent a specific experimental condition as defined by the sign matrix.

A study on target acquisition performed at the Naval Weapons Center, China Lake, California (Grossman and Whitehurst, 1976) illustrates how subject characteristics can and should be introduced into the experimental design of the screening study. Three of eleven factors in that study were subject factors.* This required a minimum of $2^3 = 8$ subjects, each having the appropriate combination of characteristics as indicated in the following sign matrix:

Subject #	Acuity(A)	Depth Perception (B)	Color Vision(C)
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

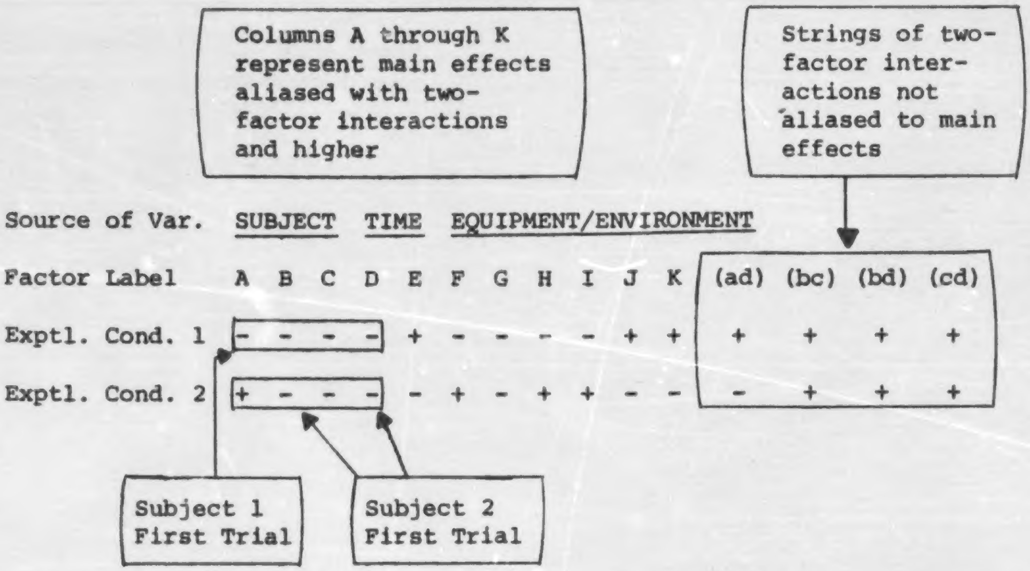
Where - represents the poor condition and + represents the good, according to specified criteria.

Each subject was tested under appropriate combinations of the eleven equipment/environment factors required to complete the 16 experimental conditions of the complete 2^{11-7} Resolution III design. For example:

---

*A fourth, labeled Experience (D) might be considered a subject characteristic but was introduced into this experimental design as a temporal factor. Each subject ran through the experiment twice. The first measurement of each condition was considered low experience and the second measurement of each was considered high experience.

70



Columns A through K represent main effects aliased with two-factor interactions and higher

Strings of two-factor interactions not aliased to main effects

Source of Var.	SUBJECT			TIME		EQUIPMENT/ENVIRONMENT								
Factor Label	A	B	C	D	E	F	G	H	I	J	K	(ad)	(bc)	(bd)	(cd)
Exptl. Cond. 1	-	-	-	-	+	-	-	-	-	+	+	+	+	+	+
Exptl. Cond. 2	+	-	-	-	-	+	-	+	+	-	-	-	+	+	+

Subject 1
First Trial

Subject 2
First Trial

Although a number of subjects are involved, each experimental condition is represented only once in an unreplicated design. If one considers subject factors equally as important as equipment factors, then no distinction need be made in the analysis of the data. If the purpose is to order the factors whatever the source, according to their relative effects on the performance of the task under investigation, then this screening design can be used.

Measuring Subject Characteristics Not in the Design

When it is not possible to vary subject parameters by systematically selecting a subject with precisely the correct combination of characteristics, then measurements should be made of the characteristics as they actually exist in the subjects who are used. If over the entire experiment the variables tend to distribute themselves relatively normally, then their effects can be estimated along with the more systematic ones using a regression analysis. One can visualize the variables laid out as terms of a polynomial to estimate performance, $\hat{y}$:

$$\hat{y} = \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_5 E + \beta_6 F \ldots.$$

where the italicized letters represent measured values of the uncontrolled variables (probably correlated among one another and the other variables) while the Roman letters represent selected levels of the controlled factors of the factorial (or fractional factorial). The β_i are the weights of each variable as determined by a regression analysis, preferably ridge regression analysis (Simon, 1975). As the correlation among variables increases, ridge regression analysis is superior to the conventional multiple regression analysis for this purpose. However, when uncontrolled variables are to be measured and analyzed along with the controlled variables in the experimental design, enough extra observations must be made to provide the degrees of freedom needed to cover the additional uncontrolled variables.* These degrees of freedom may be obtained if the basic design is replicated using a representative sample of different subjects selected in some random manner. The required number of degrees of freedom may also be obtained if the orthogonal design is analyzed first in the prescribed manner, and those factors that are definitely trivial are dropped from the analysis. Presuming the number of factors dropped is equal or greater than the number of measured variables, then there will be enough degrees of freedom available to re-analyze data to include the uncontrolled but measured variables (co-variables). While there are some dangers associated with this latter procedure, an alert investigator should be able to detect them if they arise. The odds favor the latter approach which maintains the integrity of economy in a screening study.

---

*Use of this technique need not be limited to uncontrolled subject variables, but can be applied for any type of uncontrolled variable.

SUBJECTS AS REPLICATION

Replication is the antithesis of experimental economy. In some cases, it is used unreasonably. Such is the case when an investigator replicates a fractional factorial design. If he intends to expend this additional effort collecting more data, it would be far more informative to add a different fraction to the design than it would be to replicate the original fraction. In this way, more sources of variance in aliased strings could be isolated, increasing the investigator's understanding of the situation. As Daniel (1976, p 10) says: "The most useful replication will be that which best samples the population of conditions about which E wants to make inferences. In this sense, the best replication is done under different conditions, not under the same conditions." Simon (1973, pp 19-32) reviewed the arguments psychologists frequently give for replicating, and indicated their weaknesses and alternative solutions.

Two valid reasons for "replicating" with subjects, after all other alternatives have been exhausted, are to establish inter-subject reliability and to obtain confidence intervals or fiducial limits.

Replicating for Inter-subject Reliability

An investigator never really knows if there are unwanted and unknown sources of variance affecting his experimental data. No matter how careful he may be -- and there appears to be large investigator differences in the care with which they collect experimental data (Simon, 1976b) -- an investigator should impose checks on the quality and consistency of his data. This means that when a second subject is tested on all the experimental conditions, the data from each subject should be analyzed *separately* and *compared*. This not only permits a check on the consistency of responses among homogeneous subjects as well as the assumption of homogeneity, but also helps detect distortions and outliers in the data. Some hints in this regard are discussed in the section on Analysis. The investigator may even wish to test more subjects (still making individual examinations of the results) until he builds confidence in a particular set of conclusions or discovers reasons for not accepting them.

While methods of isolating experimental from trend effects in screening studies have been described, an investigator may be as concerned with cross-over effects as he is with trends. If so, he may decide to present the conditions to several subjects in different orders in a way which will enable cross-over effects to be isolated from experimental effects (Simon, 1974, pp 27-90)*

---

*Economical designs that are robust both to trend and cross-over effects have not been worked out.

Replicating to Establish Confidence Intervals

The appropriate research strategy is to establish confidence intervals at the end of the experimental program. Once an equation containing all of the critical factors has been derived, those combinations of factor values that optimize performance or represent combinations of practical interest would be used to test a group of "truly" homogeneous subjects. Subjects can be considered homogeneous after the investigator has separated them into groups on the basis of critical subject variables and any remaining within-group subject variability is small and not readily identifiable. It's the "what's left over" after all efforts to identify the sources have been exhausted. Generally, establishing confidence intervals would be done in the operational environment where that information would be most useful.

DATA ANALYSIS

- How to calculate the criteria for deciding which factors are critical to the task under investigation and which are marginal or trivial: effects, eta squared, cumulative proportion of variance, half-normal plots.
- How to analyze subjects used to replicate the basic screening design.
- How to adjust experimental effects for trends.
- How to analyze multiple responses: graphical and statistical methods.
- How to evaluate how well first and second order regression equations fit the empirical data.
- How to analyze an incomplete screening design.

V. CALCULATING CRITERIA TO SELECT NON-TRIVIAL FACTORS

Since the purpose of the screening study is to identify those factors out of a larger candidate group which have non-trivial effects on performance, the first step of the analysis is to calculate a number of criteria which will help the experimenter make that judgment. It is appropriate at this time, before the analysis begins, to emphasize the point that there are no mechanical methods of selecting the trivial and non-trivial factors. Lest the unsophisticated investigator believe that requiring subjective decisions on the part of the investigator is unscientific and is a weakness confined to these screening studies, let him be assured that this is not the case. Evaluating the results from a screening designs study is no different from evaluating the results from an analysis of variance by hypothesis testing. Accepting or rejecting the hypothesis is done by the investigator, not the F-test (Bakan, 1967). Statistics applied to the empirical data may facilitate a decision.

SELECTION CRITERIA

Whether or not a factor is considered non-trivial will be based on the following criteria:

1. Does it have a practical effect on performance? This can be determined by calculating its effect, i.e., the mean difference between the high and the low value of that factor.

   Precaution: If the pair of values per factor in the experimental design does not cover the full range of interest, an estimated effect will not be indicative of the full strength of this factor.

2. Does the factor account for a meaningful proportion of the variance in the experiment? This is determined by calculating Eta squared, or the ratio of the sum of squares for the factor to the total sum of squares.

   Precaution: If the candidate list does not include essentially all of the critical factors affecting performance under operational conditions, then proportions obtained in the experiment will be deflated when applied to a real-world problem.

3. Does including the factor materially improve the ability to predict performance under operational conditions? This is determined by examining the cumulative proportion of variance obtained when the effects of the factors are combined.

   Precaution: If an effect is due to chance in this sample, including it will lead to poorer prediction in subsequent tests (shrinkage).

4. Could the observed effect have been due to chance? Without a source of error variance, the investigator must rely on less direct indications (i.e., internal tests) of a chance phenomenon. Examining the data using "half-normal plots" may be useful for this purpose.

   Precaution: This graphic inspection of portions of the data is still a poorly developed art.

5. Can the cumulative effects of a large number of non-critical factors be ignored?
While some factors may show only small effects, nevertheless, they have an impact on performance. If there is a large number of marginal factors, and according to the principle of maldistribution that is what we expect, we may wish to exclude them during an initial screening, but to examine them more carefully during the refinement phase of the program. Together they may improve prediction considerably.

In applying the above criteria, the investigator will temper his judgment with the cost of each decision, as well as by satisfying the interests of those who have sponsored the research. With an iterative research strategy, and decisions are constantly being tested, no decision need be final. Factors included or excluded early in the program may be excluded or included later in the program, if necessary. Generally the error in decision will occur with the marginal factors where the practical effect of an error is the smallest.

Estimating the Effects

The "effect" of a factor is the mean difference between the performances measured on the two levels or conditions of that factor.* An investigator has several methods at his disposal for estimating effects in the 2^{k-p} screening design.

---

*Some statisticians use the term "contrast," instead of "effect."

The conventional method (for psychologists) of finding the effect of Factor A, for example, in a 2^{k-p} design, would be to add up all of the performance values in all the cells associated with one condition of Factor A and to add up all of the performance values associated with the other condition of Factor A. The means of these sums would represent the mean performance on the two conditions and the difference between the two means is referred to as the "effect." This is illustrated with some fictitious data in Table 5.

As more factors are included in the screening experiment, the sign matrix can be used to facilitate the analysis. The sign matrix for a 2^{2+1}_{V} design, along with fictitious performance data, is used to show how a sign matrix is used (Table 6). For example, the effect of Factor A is estimated by summing all performance scores, obtained when the "high" (+) condition of A is being tested, e.g.:

$$4 + (-5) + 3 + 5 = 7$$

This sum is divided by 4, giving the mean performance of 1.75 for the high condition. Next, all performance scores, obtained when the "low" (-) condition of A was being tested, are summed, e.g.:

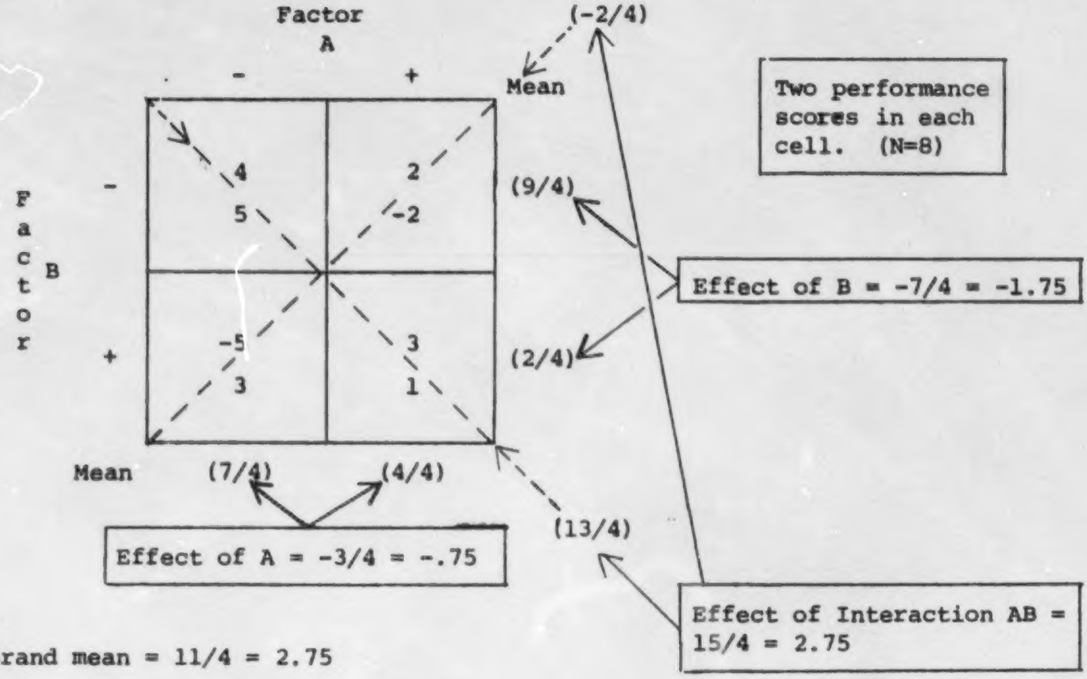
$$2 + 3 + 1 + (-2) = 4$$

This sum is divided by 4, giving a mean performance value of 1.00 for the low condition. Subtracting the low from the high gives a mean difference of 0.75, the effect of Factor A.

Similarly, the effect of the interaction AB would be obtained as follows:

$$(+4 + 3 + 1 + 5) - (+2 -5 +3 -2) = 15 \div 4 = 3.75$$

TABLE 5. CONVENTIONAL METHOD OF ANALYZING 2^k DESIGN


Factor A
−
+
Factor B
−
+
4
5
2
-2
-5
3
3
1
Mean
(7/4)
(4/4)
Mean
(-2/4)
(9/4)
(2/4)
(13/4)
Two performance scores in each cell. (N=8)
Effect of B = -7/4 = -1.75
Effect of A = -3/4 = -.75
Effect of Interaction AB = 15/4 = 2.75

Grand mean = 11/4 = 2.75

TABLE 6. SIGN MATRIX METHOD OF ANALYZING 2^k DESIGN

		Factor Sources			Performance
		A	B	AB	
Experimental Condition	(1)	−	−	+	4
	a	+	−	−	2
	b	−	+	−	-5
	ab	+	+	+	3
	ab	+	+	+	1
	b	−	+	−	3
	a	+	−	−	-2
	(1)	−	−	+	5

A	
−	+
4	2
-5	3
3	1
5	-2
7	4

Effect of A = $\frac{-7 + 4}{4}$ =

-3/4 = -.75

The divisors for the means in these cases always equal half the total number of performance values. Each column of the matrix, i.e., each source of variance, is treated in the same way.

Yates' algorithm. When the effects of a large number of factors must be estimated, using the sign matrix can become tedious and the chance of making arithmetic errors increases if a computer is not used.* Yates (1937) developed a systematic tabular method of calculating the effects of 2^k designs which is adaptable to 2^{k-p} designs including screening designs. An example of the analysis of a 2^3 design using Yates' algorithm is given in Table 7. The steps are these:

1. List the 2^k experimental conditions in the Standard Order (Column I). This Standard Order is (1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, and so forth, where after the (1) condition, a factor at a time is added, followed by all interactions between that factor and each previous factor combination before a new factor is added.

2. List each performance score adjacent to the corresponding conditions (Column II). If it will simplify

---

*Even if the calculations are done with a computer, there is a material advantage in using Yates' algorithm. Because the Standard Order is assumed (or corrected for later if the initial assumption is incorrect), the only inputs to the computer are the performance scores. No 2^k or 2^{k-p} matrix need be input, a savings in the programming and card punching requirements.

TABLE 7

YATES' ALGORITHM FOR ANALYZING A 2^k FACTORIAL

I	II					III	IV
Standard Order	Experimental Condition	Performance	1	2	Effect-total 3	Effect ÷(N/2)	Source
1)	(1)	4	6	4	11	2.75	Mean x 2
2)	a	2	-2	7	-3	-.75	A
3)	b	-5	3	6	-7	-1.75	B
4)	ab	3	4	-9	15	3.75	AB
5)	ab	5	-2	-8	3	.75	C
6)	b	-2	8	1	-15	-3.75	AC
7)	a	3	-7	10	9	2.25	BC
8)	(1)	1	-2	5	-5	-1.25	ABC

calculations, a constant can be subtracted from every score without affecting the estimation of the effects. Only the mean must be corrected by that constant amount.

3. Separate the numbers in Column II into pairs and add the two values in each pair, taking signs into consideration. List these sums in order in the upper half of Column 1.

   Next, start again at the top and _subtract_ _the_ _FIRST_ _number_ _from_ _the_ _second_ of each pair in Column B and list the differences in order in the lower half of Column 1.

4. Repeat this process to create Column 2 using the numbers in Column 1.

5. Continue to add and subtract adjacent pairs in each list to create a new list until there is a total of _k_ numbered columns for 2^k experimental conditions. In the example in Table 7, with $8 = 2^3$ conditions, there are three numbered columns (Columns 1, 2, and 3).

6. The effect for each factor is obtained by dividing the appropriate value in the _last_ numbered column, referred to as the "effect-total,"* by a value equal to half the total number of observations in the experiment (Column III).

---

*Sometimes called the "contrast-sum."

The effects thus calculated will also be listed in the Standard Order, the first value being equal to twice the mean* of all the data, the second being the effect of Factor A, the third being the effect of Factor B, the AB, C, AC, BC, ABC, D, AD, and so forth (Column IV).

When this analysis is used with screening (or other fractional factorial) designs in which the original factorial labels are changed to new screening design labels, and effects are aliased, an equivalent change must be made in the factor labels of the analysis using Yates' algorithm. Corresponding new labels must be substituted for the old labels that appear in the Standard Order in the effects column.

Daniel (1956, p 93) writes: "With N_R as large as 32, Yates' computational form may be split into two forms of size 16, using sum and differences of pairs over the last factor, instead of the original single results. This subdivision may be continued further for N_R larger than 32." This is illustrated in Table 8. The performance values (Column B) of the experimental conditions (Column A) listed in Standard Order would be divided in half, with performances associated with all low conditions of the last factor (i.e., C in this example) being analyzed with Yates' algorithm as one problem and performance associated with all high conditions of the last factor analyzed as a separate problem. Since only half the data is in each problem there will be one less column in each sub-analysis (Columns 1 and 2) than would be in the full analysis. When the effect/total values

---

*We divided by N/2 in Step 6. To get the mean, we would of course divide by N.

TABLE 8

USING A SPLIT-YATES' ALGORITHM TO ANALYZE A 2^k DESIGN

	I	II	1	2	2'	2"	3	(÷4) III
Half with C-	(1)	a	a+b	a+b+c+d =	*(A*	*A*	*A+E*	-
	a	b	c+d	b-a+d-c =	*(B*	*E*	*B+F*	A
	b	c	b-a	c+d-a-b =	*(C*	*B*	*C+G*	B
	ab	d	d-c	d-c-b+a =	*(D*	*F*	*D+H*	AB
Half with C+	c	e	e+f	e+f+g+h =	*(E*	*C*	*E-A*	C
	ac	f	g+h	f-e+h-g =	*(F*	*G*	*F-B*	AC
	bc	g	f-e	g+h-e-f =	*(G*	*D*	*G-C*	BC
	abc	h	h-g	h-g-f+e =	*(H*	*H*	*H-D*	ABC
	Original exptl. conditions in Standard Order	Performance values (symbolic)	Applying Yates' algorithm separately to each half split on low level and high level of factor C		Effect-totals for each half of total exper.	Effect-totals of each half after pairing	Overall effect-totals for complete exper.	Combined effects in Standard Order
	I	II	1	2	2'	2"	3	III

are obtained for each half (Column 2') they would be intermingled, alternating with the first effect-total in the low-condition analysis followed by the first effect-total in the high-condition analysis, and continued to alternate in this fashion until the two halves are completely paired (Column 2"). This new column is then treated to the sum-difference analysis as if it had been the next to last column, the effect-total values, of the full analysis (Column 3). Then in this example, if Column 3 is divided by N/2 = 4, we obtain the mean doubled and the effects in Standard Order.

Estimating the coefficients of the multiple regression equation. In screening designs, the equation would take the form:

$$\hat{Y} = b_oX_o + b_aX_a \ldots b_kX_k + b_{\underline{ah}}X_{\underline{ah}} + b_{\underline{ag}}X_{\underline{ag}} \ldots + b_{\underline{ab}}X_{\underline{ab}}$$

where

$\hat{Y}$ = estimated performance

b_i = coefficient for factor i, where i = a to k

X_i = term representing the level of factor i; $X_o = 1$

$b_{\underline{ij}}$ = coefficient for the string of two-factor interactions

$X_{\underline{ij}}$ = term representing the string of two-factor interactions

The regression coefficient, b_i, equals $\Sigma YX_i/\Sigma X_i^2$. However, in the basic screening design, $\Sigma X_i^2 = N$ and ΣYX_i equals the effect-total in the Yates' algorithm. Thus the regression coefficient for a multiple regression equation can be obtained as follows:

$$b = \frac{\text{Effect-total}}{N} = \frac{\text{Effect}}{2}$$

Interpreting effects data. The effects data shows the change in performance that occurs between the two levels of each factor. If these levels represent the extremes of the operational space, or the upper and lower limits of performance, then the magnitude of the effects tells us something of the practical importance of that factor for the task under consideration. Thus, it is not possible to make a meaningful interpretation of the results without fully understanding the design and its context in the real world. In Figure 5 (solid line), the effects of resolution might be quite different depending on which two levels had been selected for the conditions of the experiment:

Trivial:	AB, DE
Mild:	CD
Large:	BD

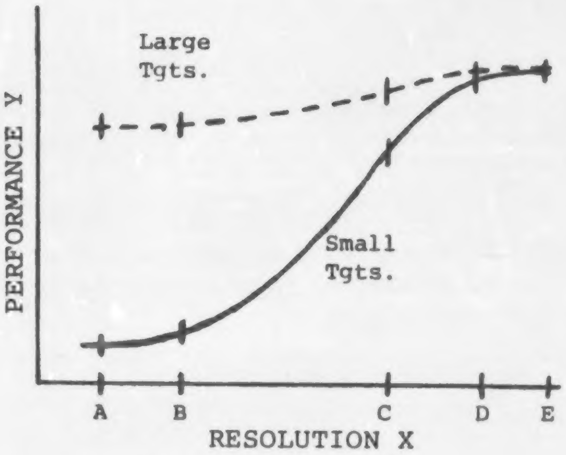


Figure 5. Illustration of How Experimental Context (i.e., task difficulty) Affects Performance

All this could change as a function of other parameters. For example, the effect between levels B and C might have been trivial if all targets had been so large that differences in resolution were inconsequential. This is illustrated by the dotted line.

When effects are evaluated, however, the interpretation must be made in the context of the operational situation rather than the experiment. It is equally important that the performance also be measured in terms of operationally relevant parameters. For example, a 2.4-second difference in the speed of reading a full-size newspaper page would probably not be an important consideration in selecting one of two styles of type. On the other hand, a 2.4-second difference might be quite critical in selecting the design of a safety switch on a nuclear reactor. The experimenter must look at the effect and decide if one that size is critical in the performance of the real-world task. If it is definitely not, then that source of variance can be excluded until new evidence negates that decision. If it is a marginal effect, other considerations involving costs and convenience will determine whether it should be excluded at this time or not. If the effect being considered represents the sum of a string of two-factor or three-factor interactions, the investigator should determine whether or not any of the larger main effects are paired in the string. If so, it is likely (though definitely not certain) that the string represents an ordinal interaction which is of secondary importance. Deciding whether a string contains an ordinal or disordinal interaction may require more data to be collected (Simon, 1973, p 116-124).

Estimating the Proportion of Variance Accounted For

For an unreplicated 2^{k-p} screening design, the variance, or mean square, can be calculated quite simply once the effects for each source have been obtained.

First the Sum of Squares for each individual source of variance is calculated as follows:

$$\text{Sum of Squares} = \frac{N\ (\text{Effect})^2}{4}$$

where N is equal to the total number of observations in the experiment. Since these designs involve factors at only two levels, each source has only one degree of freedom. Therefore, the sum of squares and variance for each effect are equal.

Eta squared. The proportion of total performance variance in the experiment, accounted for by each source of variance, is calculated as follows:

$$\text{Eta squared} = \frac{\text{Sum of squares for particular source}}{\text{Total sum of squares}}$$

Total sum of squares is obtained by summing the sums of squares for all N-1 sources of variance, including those between blocks, if any, in the experiment. The mean is not included.

Interpreting proportion of variance data. In interpreting the proportion of variance associated with a single source of variance, two things must be remembered: one, it is a relative measure and two, its importance depends on how many critical factors are included in the experiment. As a relative measure, the magnitude of an eta squared depends on the magnitude of the other sources of variance in an experiment. Since there is always an upper limit of 1.00 on the proportion of total variance that can be accounted for, a source that shows a mean difference of 30 seconds may, for example, account for 25% or 75% of the total variance depending on whether the other effects and error in the experiment are relatively large or small, enhancing or decreasing the absolute total variance, and changing the relative proportion accounted for by any single effect. With only one factor plus some random error variance, a factor may account for 90% or 10% of the total variance depending whether or not it is a "clean" experiment with a little or a lot of random error respectively. Thus, in interpreting eta squared, a source that accounts for a small proportion of variance is likely to be a non-critical source of variance, but a source that accounts for a large proportion of variance cannot per se be considered critical. It may have accounted for most of the performance variability in the experiment, but in the real world where a great many factors are likely to be operating, it will account for relatively little. It is the case of a big frog in a small pond.

The only time when a source with a large proportion-of-variance value can be considered critical -- with confidence -- is when a large number of factors has been included in the experiment and these are believed to include most of those likely to be critical under operational conditions. Other considerations in the interpretation of eta squared are discussed by Simon (1976b, pp 37-43).

Cumulative Proportion of Variance

When the sources of variance are ordered from largest to smallest according to the size of each one's effect and the proportion of variance accounted for by each is calculated, these proportions may be added, one at a time, so that as each new source is added incrementally, the cumulative proportion of the total variance accounted for by all sources of variance, both factors and interactions strings, up to that point, is indicated.

Since the sources in a screening design are independent, each cumulative proportion of variance represents the square of the multiple regression coefficient (R^2) for an equation composed of all sources included up to that point. Each new source adds some incremental amount, which may or may not be important -- which is what the investigator is trying to decide -- and which may in fact have been a chance effect for the particular sample and would not likely re-occur were the experiment repeated. As stated earlier, we have no way in the single replication screening design of directly measuring the error variance with which to test the reliability of each new term. Several indirect (or internal comparison) measures will be suggested later on.

When we stop at a particular point along the ordered continuum of sources and calculate the cumulative proportion of variance (or R^2), we are implicitly assuming -- tentatively at least -- that the remaining proportion of variance not accounted for (i.e., $1 - R^2$) is error. This estimate of error might be used to determine at what point the addition of another term (source of variance) results in a drop in the population R^2, which is estimated by applying certain correction factors to the sample R^2. Quite obviously, the R^2 value for the sample must increase toward 1.00 as more sources of variance are added, but the population R^2 reaches a point where instead of increasing as more sources are added, will decrease. This could be used as a clue as to where to stop adding more sources.

While there are a number of formulas to calculate "shrinkage" (Kerlinger and Pedhazur, 1973; Url and Eisenberg 1970), the following one is probably as effective as any for our present purpose and is simple to use:

$$\hat{R}^2 = 1 - \frac{N - 1}{N - k - 1} (1 - R^2)$$

where $\hat{R}$ is the population multiple correlation corrected for shrinkage

R is the uncorrected sample multiple correlation for the k factors

N is the total number of observations

k is the number of factors (or sources of variance) included in the equation

At some point, as k increases, $\hat{R}^2$ will begin to decrease. This is the maximum number of sources that should be considered. In practice, however, this formula gives an *overestimation*, and the k sources are probably *too many* to include.

Because a successful experiment should account for most* of the performance variance, there is often a tendency to want to include more sources of variance than are probably necessary. Still, the final decision of what to include or not will be made more on the basis of practical considerations and the dangers of an erroneous decision than on the results of a statistical test. The decision is made more difficult, however, when we look at the cumulative proportion than at the proportion accounted for by an individual term factor. For in individual cases, we may see a small value, e.g., a proportion of .01, and decide that even if it were a real effect, it is marginal and if we omit it erroneously it is not going to be too critical. On the other hand, we might hesitate dropping ten or fifteen effects that individually might each account for a probability value of .01 or less, since cumulatively they might, for example, account for .10 to .20 of the total variance. Luckily, the problem is easier to resolve in the screening phase when we are only asking whether a particular factor should be included in

---

*Without more experience, what proportion should be accounted for by a screening design cannot be stated with any degree of confidence. Still, as a personal guess, if we started with a 30-factor study (and an astute experimenter), one ought not to be happy unless one accounts for *more than* .80 of the variance in the experiment with *real* effects.

subsequent studies than it would be in the refinement phase near the end of the research problem, when even small amounts (as long as they are real effects) should not be ignored. But in the screening phase, if even a series of factors shows a sizeable cumulative effect, if they have been preceded by a great many interaction strings each with meager effects, and occur in the second half of the ordered sources of variance, it is unlikely that any effect will be critical.

Reverse Yates' algorithm. Daniel (1976, p 73) examines the cumulative proportion of variance one step at a time using a reverse Yates' algorithm as a computational aid. Beginning after a reasonable number of terms has been included in the cumulative proportion, he calculates the predicted value at each experimental data-collection point in the design and compares it with the empirically obtained value. Calculating the predicted values could be done using the regression equation, however, Daniel's application of the reverse Yates' algorithm is the same as for the forward Yates' with the following exceptions:

1. Begin by writing the effects in the Standard Order, but inverted.
2. Read off the estimated values at the end of the procedure with the conditions in an inverted Standard Order.

If one begins this reverse analysis with the effects, then the values in what would ordinarily be the effects-total column must be divided by N/2 to get the estimated performance values. However, if instead of beginning with the effects one begins with the regression coefficients, then no division is required. The values in the effect-totals position of the reverse Yates' are the estimated performance values. For example:

FORWARD YATES'

Exptl. Cond.	Perf. (y)	1	2	(÷2) Effect	Source
(1)	3	11	18	9	2(M)*
a	8	7	2	1	A
b	5	5	-4	-2	B
ab	2	-3	-8	-4	AB

REVERSE YATES'

Source	Effect	1	2	(÷2) Est. Perf. $\hat{y}$	Exptl. Cond.
AB	-4	-6	4	2	ab
B	-2	10	10	5	b
A	1	2	16	8	a
2(M)*	9	8	6	3	(1)

The proportions of variance accounted for by A, B, and AB are .048, .190, and .762, respectively. If these were ordered from largest to smallest, the cumulative proportion would be:

AB	.762
AB + B	.952
AB + B + A	1.000

In this simplified example, Daniel might propose to find out what the estimated performance would be for each condition if we assume that A is actually zero for all practical purposes. Using the reverse Yates' he would get:

---

*Value in Effect column is twice the value of the mean.

Source	1	2	3	(÷2)	$\hat{y}$	y	$(y - \hat{y})$	Exptl. Cond.
AB	-4	-6	3		1.5	2	.5	ab
B	-2	9	11		5.5	5	-.5	b
A	0	2	15		7.5	8	.5	a
(M)	9	9	7		3.5	3	-.5	(1)

where $\hat{y}$ is the estimated value and y is the obtained one. He would test to see if the residual, $(y-\hat{y})$, could be tolerated or not, and thereby decide whether the dropped variable, A, can be excluded or not. In this artificial example there was no mean difference and no source of error variance, so no significance test would be meaningful. In the case of larger designs, however, this is yet another tool to help the investigator judge whether to include or omit a source of variance.

Daniel also uses this calculation to discover whether there are distortions in the data and whether transformations could be used to simplify the model. In particular, he plotted the residuals (i.e., the $\hat{y} - y$) against their corresponding performance (y) values as proposed by Anscombe and Tukey (1963), and also their distribution on a normal cumulative distribution grid. He next searched these for patterns that would be indicative of distortions in the data. While the study of residual patterns is an important part of the data analysis process, no further discussion on this topic will be given in this report. It is described in detail in Daniel's (1976, pp 71) book.

Interpreting the cumulative effects of non-critical factors. There is something disconcerting when it is discovered that the non-critical factors, (i.e., the ones that individually account for only a small proportion of the variance) in combination, account for a large chunk, perhaps .30, of the total variance. That is a great deal of unexplained variance and it may cause an investigator to think that possibly some of the non-critical factors may be marginal ones of minor but practical interest. He may wish to examine these non-critical factors in order to decide which he still believes are trivial and which might be considered real but "marginal." Some considerations in this regard are listed below:

1. The small effect may in fact be trivial, a chance perturbation. It is unlikely to be found on subsequent tests.

2. A noticeable effect might be due to error, an infrequent and intermittent disturbance in a few cells, affecting by chance a particular effect or two. For example, momentary losses of attention on the part of the subject, an irrelevant but intermittent occurence in the environment, or erroneous settings of the simulation equipment. The momentary effects are large, but are averaged down in the analysis. An examination of the raw data or a half-normal plot may reveal this.

3. The effect may reflect an unexpected confounding with some concomitant, systematic, but irrelevant source of variance. This might not occur if the study were repeated and can often be avoided with better planning during the problem definition phase. The size of the observed effect may be distorted due to the confounding effect a) inflating a factor's otherwise trivial effect, or b) deflating an important factor's effect.

4. The effect may be reliable, but small. More measures will be required to see if the effect occurs consistently. It might have been larger had a different part of the operational space been included in the experiment.*

The investigator, faced with the decision to include or exclude the marginal factors, realizes that:

a) If he includes a marginal factor, he adds to the expense in subsequent efforts that must allot more observations to study that factor, more time to change the factor during the experiment, and more money and manpower to build and maintain the factor into the simulation. If there are no major expenses associated with the inclusion of a marginal factor, then it might as well be included. If it is a wrong decision to include it, i.e., if it is not a reliable effect, it can be deleted later.

b) If he excludes a marginal factor, he will be able to reduce the size of subsequent studies and possibly their costs, but if it is a real effect, his ability to predict will be reduced. Since it is a marginal factor, the error -- to exclude or include incorrectly -- will be relatively small. The balance arises when the fewest factors account for most of the variance in the experiment. By building a framework --

---

*This does not mean that one should artificially extend the boundaries of a factor just to get a larger (or more significant) effect. We wish to order effects by their size within a particular operational space.

> a response surface -- involving these critical variables, the marginal factors can be introduced into it at a later phase of the research program -- to refine the original equation -- when they can be investigated more thoroughly and with more precision than if they had been entered early during the screening phase.

Costs, interest, probable impact, difficulty, realism, reliability and so forth, are all weighed in the inclusion/exclusion decision regarding marginal factors.

Half-normal Plots (Daniel)

When a large number of effects are being investigated, the largest effects can be several times larger than the average even when no effect is real. In an experiment with 31 effects, the size of the largest effect could be 2.4 times larger than the average size when in fact the difference was due only to chance. Using the traditional .05 significance level in such an experiment would cause unreal effects to be judged real in over half of all experiments done (Daniel, 1959, p 312). While an examination of mean differences and eta squared values can help the investigator avoid trivial effects, these measurements do not provide sufficient data to protect the investigator from including effects which may appear to be non-trivial but which are, in fact, chance deviations.

Conventionally, t- or F-tests are used to protect the investigator from overenthusiasm regarding a large effect. Since economy is of paramount importance and replication is avoided in the screening design, there is no internal data with which to estimate the error variance needed for the significance test. In the physical sciences, error variance can be estimated from the results of other experiments studying the same problems; this would be foolhardy to try in psychology. Psychologists, who run unreplicated factorial designs, often use higher-order interactions -- generally more than three factors -- to estimate the error variance. This is done on the assumption that the effects of these interactions are negligible. However, in screening designs this is not possible since all higher-order interactions are confounded with main and two-factor interaction effects.

Of course, if any strings of two- and three-factor interactions are trivial, they can be used to estimate error. But here we are faced with an enigma since we have no error term to test whether these interaction strings are trivial. Birnbaum (1959) suggests that instead of assuming that certain interactions are zero, an inference procedure be used which assumes that a specified number of effects out of a total number are non-zero. However, he develops the mathematics only for the case where it is necessary to discover whether one effect out of many is real or not. He concluded that his statistic in that situation would be about as sensitive in detecting one real effect among thirty-one effects (if one real one were present) as traditional multiple t-tests were capable of detecting one among 15 possible effects with ten degrees of freedom for error, or one from 31 possible effects with over 20 degrees of freedom for error. We are, of course, more interested in those situations where more than one source of variance is likely to be critical.

Daniel (1956; 1959) developed a graphic method (corresponding in principle to Birnbaum's statistic) for examining the results from an unreplicated design to help judge the reality of the largest main effects and interactions, and to indicate the presence of unruly data. His method is to graphically compare the empirically derived cumulative distribution of the effects with a cumulative distribution derived from a normally distributed population. To do this, the results from the experiment are plotted on "half-normal grid" paper.

Preparing half-normal grids. The steps to produce a half-normal grid are as follows:

1. Obtain a sheet of Probability Scale graph paper. This paper is produced commercially (e.g., Keuffel and Esser Co., #358-23). On this paper, a graph of the theoretical normal distribution would be a straight line through the origin.

2. Use that portion of the grid that begins with the probability, P, of 50 and goes up to a value greater than 99. (Note: These "probability" values, of course, are multiplied by 100 to eliminate having to print the decimal.)

3. Rescale the graph paper with new probability values, P', calculated from the old values, P, where
   $$P' = 2P - 100.$$
   For example, P = 70, and P' = 2 x 70 - 100 = 40.

4. Locate the P' along the ordinate of the grid where each ordered effect (i.e., ordered contrast) must lie. A different set of values is required for each analysis in which the total number of effects is

different. The equation to find the P' value for each particular rank is:

$$P' = [(R - 0.5)/(N-1)] \text{ X } 100^{*} =$$

where R is the rank of the ordered effect and (N-1) is the number of effects that will be plotted; it is also the total degrees of freedom with N observations.

For example, in a 2^{16-11}_{IV} screening design, there are 31 effects to be plotted. The largest effect, ranked 31, would be plotted at $P' = [(31 - 0.5)/31] \text{ X } 100 = 98.39$. The effect tenth from the top, rank 22, would be plotted at $P' = [(22 - 0.5)/31] \text{ X } 100 = 69.35$. For a 2^{32-26}_{IV} screening design, with 63 effects, the effect ranked 22 would be plotted at $P' = [(22 - 0.5)/63] \text{ X } 100 = 34.13$. The P and P' values for all ranks of designs with 15, 31, and 63 degrees of freedom (and effects) are given in Table 9. P values are probabilities (X 100) for each rank plotted on normal probability grids. P' values are the corresponding probabilities (X 100) plotted on half-normal grids. An example of a 31-effect grid is shown in Figure 6.

---

*To determine the standard score, z, of each rank position on a unit normal curve (where the N and standard deviation are both assumed to be 1) we may refer to any normal distribution table such as Beyer (1966, p 117) and look up the P -- not the P' -- value (÷ 100) associated with that rank. For example, in the above illustration

If $P' = 40$, $P = 70$, then $z = .52$.

Z-values can be used to determine the height of each rank position above zero on the ordinate of a half-normal grid which could be drawn directly rather than by extracting them from a plot on normal probability paper. The z-value will also be useful later in this paper when Zahn's work is discussed.

TABLE 9

PROBABILITY* VALUES FOR CONSTRUCTING HALF-NORMAL GRIDS**

d.f. = 15			d.f. = 31			d.f. = 63					
Rank	P'	P	Rank	P'	P	Rank	P'	P	Rank	P'	P
15	96.67	98.33	31	98.39	99.19	63	99.21	99.60			
14	90.00	95.00	30	95.16	97.58	62	97.62	98.81	31	48.41	74.21
13	83.33	91.67	29	91.94	95.97	61	96.03	98.02	30	46.83	73.41
12	76.67	88.33	28	88.71	94.35	60	94.44	97.22	29	45.24	72.62
11	70.00	85.00	27	85.48	92.74	59	92.86	96.43	28	43.65	71.83
10	63.33	81.67	26	82.26	91.13	58	91.27	95.63	27	42.06	71.03
9	56.67	78.33	25	79.03	89.52	57	89.68	94.84	26	40.48	70.24
8	50.00	75.00	24	75.81	87.90	56	88.09	94.05	25	38.89	69.44
7	43.33	71.67	23	72.58	86.29	55	86.51	93.25	24	37.30	68.65
6	36.67	68.33	22	69.35	84.68	54	84.92	92.46	23	35.71	67.86
5	30.00	65.00	21	66.13	83.06	53	83.33	91.67	22	34.13	67.06
4	23.33	61.67	20	62.90	81.45	52	81.75	90.87	21	32.54	66.27
3	16.67	58.33	19	59.68	79.84	51	80.16	90.08	20	30.95	65.48
2	10.00	55.00	18	56.45	78.23	50	78.57	89.29	19	29.37	64.68
1	3.33	51.67	17	53.23	76.61	49	76.98	88.49	18	27.78	63.89
0	0	50.00	16	50.00	75.00	48	75.40	87.70	17	26.19	63.10
			15	46.78	73.39	47	73.81	86.90	16	24.60	62.30
			14	43.55	71.77	46	72.22	86.11	15	23.02	61.51
			13	40.32	70.16	45	70.63	85.32	14	21.43	60.71
			12	37.00	68.55	44	69.05	84.52	13	19.84	59.92
			11	33.87	66.94	43	67.46	83.73	12	18.25	59.13
			10	30.65	65.32	42	65.87	82.94	11	16.67	58.33
			9	27.42	63.71	41	64.29	82.14	10	15.08	57.54
			8	24.19	62.10	40	62.70	81.35	9	13.49	56.75
			7	20.97	60.48	39	61.11	80.56	8	11.90	55.95
			6	17.74	58.87	38	59.52	79.76	7	10.32	55.16
			5	14.52	57.26	37	57.94	78.97	6	8.73	54.37
			4	11.29	55.65	36	56.35	78.17	5	7.14	53.57
			3	8.06	54.03	35	54.76	77.38	4	5.56	52.78
			2	4.84	52.42	34	53.17	76.59	3	3.97	51.98
			1	1.61	50.81	33	51.59	75.79	2	2.38	51.19
			0	0	50.00	32	50.00	75.00	1	.79	50.40
									0	0	50.00

---

*P values are probabilities (X 100) to be used on normal probability grids. Adjacent P' values are probabilities (X 100) at the same rank when half-normal probability grid is used.

**If normal probability paper is not available, grids may be constructed directly by finding the z-score equivalent to the P-value (÷ 100) for each rank and using it to measure off the distance on the ordinate scale. Z-scores can be found in most normal distribution tables, e.g., Beyer (1966, p 117).

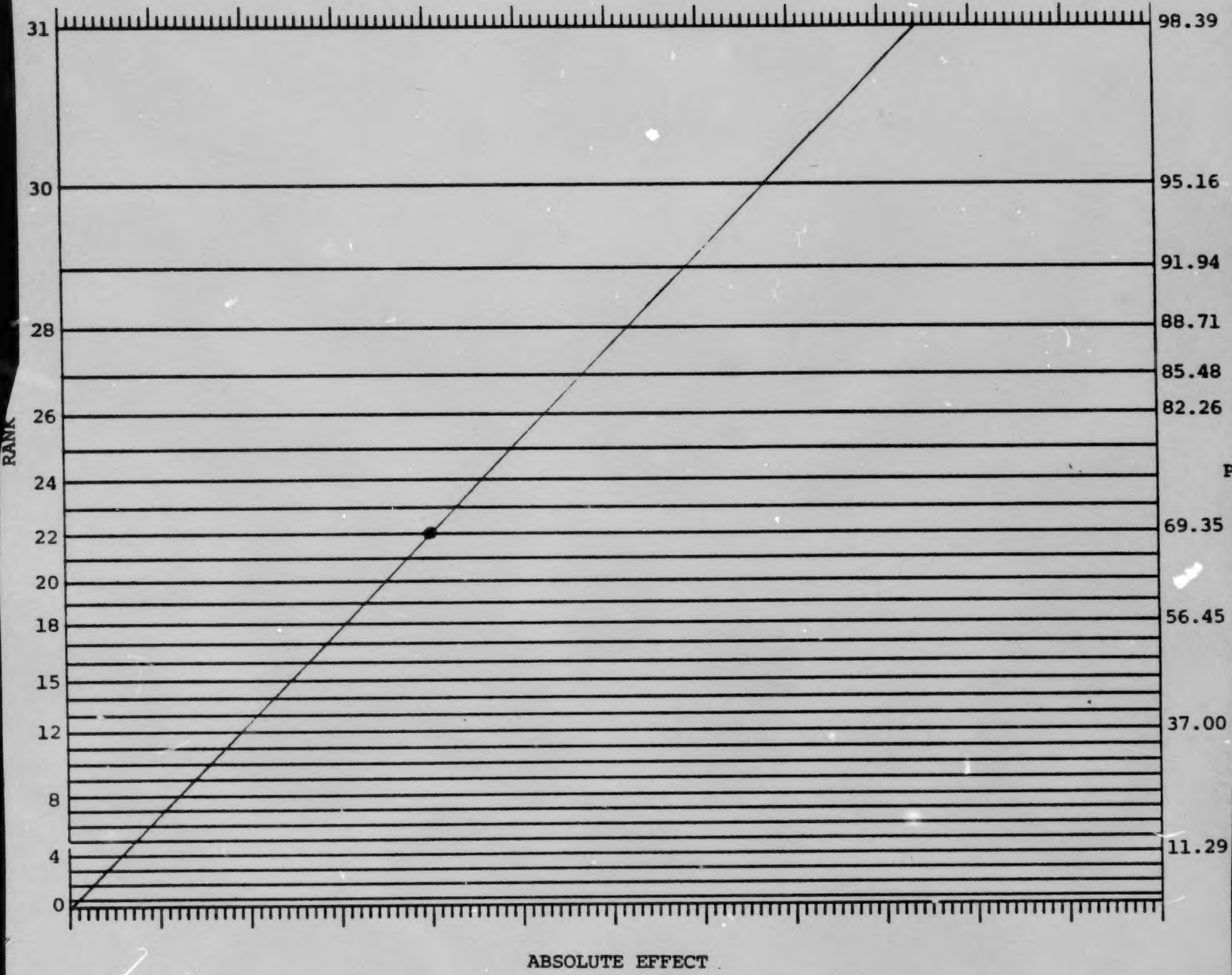


Figure 6. Half-normal Grids for Plotting Thirty-one Absolute Effects

5. Write a scale along the abscissa of the grid that covers the range of absolute* values of the effects.

Plotting the data. The absolute effects obtained from the analysis of the experimental data are ordered from largest to smallest and given the ranks from (N-1) to 1, respectively. The coordinates of a point representing the largest effect -- ignoring signs -- would be where the P' for the highest rank (along the ordinate) and the proper absolute value (along the abscissa) intersect. Each subsequent effect is plotted on the line of its appropriate rank. Daniel (1959, p 314) suggests that it is not necessary to plot every one of the smaller effects at the lower ranks since they tend to be correlated. The mean is not plotted but block differences and higher-order interaction strings, if they exist, are.

Interpreting half-normal plots. If none of the effects in the experiment are real, that is, if the sizes of the effects are no greater than might be expected to occur by chance, the standard deviation of the values would be approximated by value of the effect at the rank order nearest to the P = 68.3 quantile. In other words, the standard deviation would equal the value of the effects, X_R, when R = .683(N-1)-0.5. For 15, 31, and 63 degrees of freedom (or N-1) this would be the value at rank positions 11, 22, and 44, respectively. Under the null hypothesis, therefore, the plotted points would theoretically approximate a straight line

---

[*] The effects are ordered disregarding signs.

through the origin and the point made by the rank at the 68.3 quantile. This straight line, the "chance" line, is the cumulative distribution of a normal curve (the classic S-shaped curve) as it would appear when plotted on probability paper. Daniel (1959, p 316) plotted ten samples of 31 effects from purely random data. While the average of the ten approximates a straight line very well, individually they wander about the line in an irregular pattern, though not enough to be misinterpreted as being real effects.

In practice, since some effects may be real, we do not know exactly where the slope of the line should be. Instead, we allow the data to determine where the straight line will lie. It would be drawn by eye through points representing the smaller half of the ranked data. Ordinarily these should go through the origin, but occasionally may not. The farther the larger effects deviate to the right of this line, the more probable it is that they did not occur by chance and are in fact reliable effects.

Interpretation tactics. Krane (1963) who adapted the use of half-normal plots to multi-level factorial experiments, suggests an iterative approach to the selection of the real effects. He examines the largest point first to see if it lies far enough to the right of the line to be judged real, and if so, removes it and replots the remaining effects and again decides if the largest of the remaining effects deviates far enough from the straight line to be judged real. This continues until he no longer believes that an effect is real.

In practice, he may make a crude test of a number of the largest effects by drawing a vertical line from the point where the horizontal line on which the largest rank is located intersects the empirically constructed cumulative distribution line. He then considers only those effects lying to the right of the vertical line. Next, he replots the effects after having eliminated those largest effects already judged to be real, draws a new line and again judges whether effects to the right are real.

In the replotting, since there are fewer cases each time, the position of the rank order-lines on the P' scale must change. For example, the 31st line in Figure 6 is at P = 99.19 on the full-normal probability scale or P' = 98.39 on the half-normal plot. These values can be found in Table 9 of the effect on the 31st rank is considered real and is removed, and the remaining 30 values are replotted, the probability position of the 30th rank is now based on an (N-1)=30 rather than 31. Therefore, it cannot be plotted on the original half-normal grid in Figure 6. The new P' for each rank must be replotted using the equation:

$$P' = [(R - 0.5)/(N-1)] \times 100$$

as was done before. Or, if it is apparent that the first four largest effects can be removed, then a new P' value for the rank 27 effect would have to be calculated. P' would be 98.15.

Since a special grid has not been prepared for any size other than 31, the reader can make his own by marking off the correct grid on the upper half of the normal probability paper. In this case, he would have to work backwards in his calculations, first determining what the P' value would be for a particular rank and a particular (N-1), and then finding that position on the half of the normal probability paper at P, where P = 0.5(P' + 100).

To facilitate this effort however, calculations of P' and P for the first four largest ranks for values of (N-1) from 63 down to 4 is given in Appendix V. For example, if (N-1) equals 27, then from Appendix V we would plot on one half of a piece of full-normal probability paper the first four largest ranks at the following positions:

Rank	P
27	99.07
26	97.22
25	95.37
24	93.52

and assign the new P' = 2P-100.

Detecting defective values. Krane (1963, p 284-285) discusses Daniel's 1966 conclusions regarding the use of half-normal plots to detect defective values. These are cited here briefly to inform the reader who may be interested in pursuing this form of analysis on his own but for whom Krane's paper may not be readily available. Krane noted that the half-normal plot of an experiment involving a number of small but real interactions may appear very similar to the results induced by plot-splitting, because "split plot error contrasts invariably contain a relatively larger number of the higher order contrasts." He added: "Our practice is generally to employ a split plot analysis only when knowledge of the experimental techniques indicate its propriety." Krane also noted that because his analysis was usually based on transformed data, he seldom experienced the downward convexity of half-normal plots that Daniel, in his 1966 paper, believed indicated the presence of an antilognormal distribution of error. Krane pointed out, on the other hand, that "the removal of a moderate number of points representing apparently real effects often results in a downward convexity of the upper portion of the plot. We generally attribute this appearance to the inadvertent removal of one or more points representing error contrasts, for the results look very much like the plot of a normal distribution with truncated upper tail."

Daniel, in 1959, felt that half-normal plots could be used to detect defective values in the data. By the time he had published his book in 1976, he no longer believed that to be the case. In his book, Daniel (1976, p 149) felt that "the signed contrasts in standard order have more information in them than do the unsigned contrasts ordered by magnitude." He spends a good part of his book showing how residual analysis can be used to detect distorted experimental values. This should be an important part of the analysis of any experimental data and can be of particular value in studies employing economical multifactor designs with minimum replication. Anscombe and Tukey (1963) also treat the subject of residual analysis. This topic will not be treated in this report. Both of Daniel's books (1976; Daniel and Wood, 1971) are recommended reading for anyone analyzing applied experimental data. Unlike the authors of many textbooks on statistics, Daniel discusses and deals with the interpretation problem from a practical point of view based on years of experience.

Standardized Half-Normal Plot (Zahn)[*]

If Daniel had proposed no more than the foregoing discussion of half-normal plots, he would have made a major contribution to the analysis and interpretation of unreplicated screening design data. At the least, this type of plot warns the user that large effects might in fact be due to chance. At the most, in this computer age, it encourages the investigator to engage in that almost forgotten art of studying his data directly. But Daniel did not stop there. Instead he proposed the concept of a "standardized" half-normal plot (Daniel, 1959, p 322).

---

[*] Just when this report was ready to go to press, the papers by Zahn (1975a, 1975b) were discovered. Zahn's work (continued on next page)

Daniel proposed that a scale-free, standardized half-normal grid be used on which fixed limits could be placed to identify how far from the line deviant effects must be to have a specified probability of being a real, rather than a chance, effect. The advantage of this plan is that it facilitates comparisons among sets of data using different criteria. Furthermore, it serves as a graphic test of statistical significance, alerting the investigator to the possibility that he might be making Type I errors.

In the standardized version, Daniel's premise was that with no real effects present in the data, the standardized values of the absolute effects, when plotted on a half-normal grid, would lie along a straight line through the origin and the coordinate of the ordered effect at the rank having the value approximating the standard deviation of the data. The standardized values are obtained by dividing the absolute effects by the estimated standard deviation. Daniel estimated the standard deviation to be the value of the effect at rank, $R = .683\ Y + 0.5$ (with Y = the largest possible rank for the set of data). For data involving

---

points out flaws in Daniel's method of producing "standardized half-normal plots." Since it is believed that half-normal plots are powerful tools for interpreting unreplicated screening data, the original discussion regarding Daniel's method was removed from this report and this brief notation regarding Zahn's work was introduced in its place. The reader is encouraged to read Zahn's original papers and to use his version of the "standardized half-normal plots."

15, 31, and 63 effects, the standard deviations would be approximated by the values at ranks 11, 22, and 44, respectively, when no real effects are present in the data.

Based on the theoretical work by Birnbaum (1959), Daniel (1959, p 322) provides the data for calculating probability guidelines -- "guardrails" -- which indicate the limits above the "chance" line at which points may fall, purely by chance, a specified proportion of the time. This is a form of graphic significance test.

Zahn (1975a, 1975b) recently proposed modifications to Daniel's version of the standardized half-normal plot. He notes a minor flaw in the plotting positions and a major flaw in the method of calculating the guardrails for the standardized half-normal plots. Zahn describes two versions -- X and S -- of his own, but based on an empirical study, he concludes that his version S is the superior one (Zahn, 1975b, p 210). The difference is primarily in the way the standard deviation is calculated.

Zahn (1975a) proposes these changes in Daniel's approach to standardized half-normal plots. Two minor changes are:

1. Reorient the position of the grid so that effects are on the ordinate axis and the rank orders are on the abscissa axis. This corresponds, he felt, "to the usual regression analysis graph on which the random variable is plotted as the ordinate" (p 191). He also suggests using the raw effect values rather than the standardized score be used.

2. Make minor changes in the plotting positions (i.e., the z-values of P and P') on Daniel's grid since the standardized effects that Daniel uses are not actually half-normally distributed. Zahn (p 192-192) recommends minor changes when there are 15 effects and none when there are 20 or more effects to be plotted.

For n = 15, the ranks and Daniel's z-values are shown below along with Zahn's (1975, p 192, Table 2) recommended z-values for the new plotting positions:

Rank	Daniel's z	Zahn's z
15	2.12	2.050
14	1.64	1.626
13	1.39	1.376
12	1.19	1.191
11	1.04	1.040
10	.90	.910
9	.78	.794
8	.67	.688
7	.57	.589
6	.48	.496
5	.39	.408
4	.30	.322
3	.21	.239
2	.13	.158
1	.04	.079

P and P' values associated with Daniel's z's (d.f. = 15) can be found in Table 9, this report. These values would shift for Zahn's z. However, given the z-values, there is no reason to obtain the probability values.

The two major changes are:

1. Daniel makes his initial estimate of the standard deviation of the data on the basis of a single value, the effect at the rank position closest to P' = 68.3. For more stability, Zahn proposes, in his version S, to use a value based on the slope of the ordinary least squares regression line through the origin of the standardized half-normal grid, and fitted to the points of the smallest non-real ("error contrasts") effects, i.e., from the lowest rank, 1, up to rank a, where a equals [0.683(n + 1)].

   The estimated standard deviation so defined is:

   $$\tilde{\sigma}_{(a,n)} = \sum_{i-1}^{a} x_1 z_{ir} \Big/ \sum_{i=1}^{a} z_{ir}^2 \qquad a \leq r$$

   where

   a = 0.683(n + 1) = number of effects to be fitted

   r = largest rank

   x_i = absolute effect at rank i

   z_{ir} = standard score of rank i on unit normal probability curve (see footnote, page 86, this paper)

2. Zahn (p 195) proposes different criteria for determining the guardrails and therefore computes new guardrails. The guardrails represent the distance above the "chance" line, (i.e., the line through the smallest non-real effects) at which different probabilities of making a Type I error would occur if effects plotted above those guardrails were hypothesized as real. Specifically, Daniel's

approach failed to take into account the fact that in the single experiment we are trying to estimate whether a family of effects is significant. The probability error rate (PER) is the probability that there is at least one false positive in the family of statements. Daniel's guardrails have a valid PER only if no real contrasts are present. They were appropriate for detecting one false positive. In screening designs, we expect more and thus we would want to employ a different PER. For example, if we wish to have the Type I error rate for k = 9 real effects to be $\alpha = .05$, then the guardrail beyond which significant effects would be located on the grid would have to have a probability error rate of:

$$PER = 1 - (1 = \alpha)^k = .37$$

Zahn (1975a) uses rather elaborate statistics to calculate the guardrails for his version X (p 196) and an empirical Monte Carlo sampling study to determine the guardrails for his version S (p 197). He does provide the critical values by which new guardrails (for PER = 0.05, 0.20, and 0.40) can be plotted for N = 15 for his version S model and N = 15, 31, and 63 for his version X model. These are provided in Appendix VIII.

Zahn (1977) stated that he had done little with this work since the papers were published. As far as he knew, no one had determined critical values for N = 31 or 63 for his version S model. He suggested that the guardrails for version X might be used instead, along with the more reliable version S estimate of the standard deviation, as long as the investigator realizes that version S requires slightly larger effects than version X for the same significance level. The differences for N = 15 can be observed by comparing the values in his Table 5 and Table 7 (also reprinted in Appendix VIII of this paper) or by studying the plots, shown in his Figures 4 a and 9.

In the behavioral sciences, one can use this approach, but must beware of assuming that very precise judgments can be made. For example, in applying half-normal plots to screening designs, it is not certain that the distribution of effects (representing values from aliased sources) is necessarily the same as that of a full factorial with the same number of effects. Also, the guardrails cited here are calculated based on the assumption that a specific number of effects might be real. Thus, the critical values for plotting guardrails can vary considerably depending on the assumptions of the investigator (or the model employed in the calculation). If we do not take these mathematically precise values too seriously, we can make effective use of the half-normal plots.

These plotted values are only one of a number of criteria to be used for screening and selecting the most important variables for future study. The half-normal plots provide a check on an investigator overenthusiastically declaring effects to be real when they might have been chance. Whether the probability of the Type I error is precisely 0.40 or 0.30 is not critical in this case. Used judiciously -- and we do need more experience in using them in behavioral research -- these half-normal plots can be expected to be extremely useful evaluative tools.

USING ORDERED DISTANCES WITH MULTIPLE RESPONSE DATA

Wilk and Gnanadesikan (1961) propose a method of graphical analysis using ordered distances which represent a generalization and extension of half-normal plotting. This will be discussed later in this report. Gnanadesikan (1963) illustrates how these techniques might be used. His comments regarding the use of these "internal comparison procedures" are important from the point of view of research strategy and worth noting here. He said (p 22-23):

> *While formal procedures, with formal or informal interpretations, are useful in their own way, yet, as anyone who uses statistics learns rapidly, they do not satisfy all needs. It is neither usual nor productive to think that the real insights into data are gained by posing a few questions in terms of a few parameters and by seeking for their answers through the use of certain formal techniques. Statistical procedures, with or without a formal probabilistic framework, which are aids, in a sense, to "allowing the data to analyze themselves" are valuable tools in gaining insights into the structure of data . . . .*
>
> *Informal procedures, with their chief purpose of serving as aids to learning from data and, in a sense, unhampered by considerations of probability statements, should guide and stimulate the experimenter into partitioning the data, and studying the partitions separately, both with respect to the treatment structure and with respect to the response structure in the problem. Also, informal procedures should depend on prior as well as posterior (after seeing the data) considerations and judgment.*

Perhaps the main advantage of a tool such as half-normal plotting is that it encourages the investigator to leave his computer outputs and immerse himself in his data.

VALIDATION TEST

Wilburn (1963, p 23) proposes a validation test on the final selection of critical factors (and noteworthy interactions) to ascertain that no large distortions occurred in the actual responses that could have seriously altered the mean effects. He writes: "The procedure used was to determine the standard error of the individual observed responses by analyzing the thirty-one mean effects. A second standard error, for the difference between observed and predicted responses, was then obtained with the predicted responses based on the assumptions that all mean effects other than those for [the critical factors] were indeed zero. If the two standard errors would then be equivalent, both the total experiment and the conclusions derived from it would be proved valid."

The first standard error is estimated by ordering all of the effects of the sources of variance judged to be non-critical and using the value at the rank position R for which P' is most nearly 0.683, obtained from the equation:

$$[R = 0.683 (N-1) + .5$$

This would mean, for example, that the effect at rank 16 would serve as a rough estimate of the standard error of 23 sources, all considered non-critical.*

$$[R = 0.683 (23 - 1) + 0.5 \approx 16]$$

The second standard error is calculated as follows. First, do a reverse Yates' algorithm computation on the calculated effects after making the effects of all non-critical sources equal to zero. The answers so obtained are the "predicted" responses. Second, subtract the predicted response from the actual, observed response for each condition. Third, rank order these differences including signs. Fourth, plot them on normal probability paper $[P = (R - .05)/(N-1)]$ including signs. The difference value scale is along the ordinate; the probability (P) value scale is along the abscissa. Fifth, draw a line through the plots approximating the least squares fit. Sixth, determine the vertical distance between the .50 and the .84 P values.

---

[*] The rank nearest to P =0.683 for all cases of n from 63 down to four are given in Appendix V.

This distance read from the ordinate scale represents the standard error of these differences between predicted and observed responses. If the two standard errors are essentially equivalent, this is sufficient, Wilburn claims, to accept the experiment as being valid. (Note: Obviously, it is "valid" only insofar as the mathematics is concerned. Validity of simulation, representativeness of the subjects and task, and other features determine ultimate validity).

NUMERICAL EXAMPLE OF A SCREENING STUDY ANALYSIS

An experiment was performed at the U. S. Naval Weapons Station, China Lake, California that may represent the first attempt on the part of engineering psychologists to employ a saturated fractional factorial and foldover design for screening purposes. (Grossman and Whitehurst, 1976). In this study the effects of eleven factors on the location and identification of targets in a simulated terrain model were investigated to ascertain their relative importance in that task and to generate curves to indicate how performance varied as a function of the more important effects.

The eleven factors that were investigated are listed in Table 10. These factors could be divided into three classes depending on whether they were subject, time, or environment related. How the investigators handled the subject-related factors within the experimental design was discussed earlier in the section on the design of screening experiments. While the investigators were primarily interested in the effects of the single factor, Visual Acuity, on target acquisition, the use of this multifactor plan illustrates how a much more generalizable data base can be achieved with this approach than had acuity been studied alone.

TABLE 10

THE ELEVEN FACTORS AND THE TWO LEVELS ASSIGNED EACH FACTOR

	Factors	Levels	
		-	+
A.	Visual Acuity	20/40	20/20
B.	Depth Perception	Poor	Good
C.	Color Vision	Deficient	Good
D.	Experience	2 Trials	14 Trials
E.	Slant Range	1600 m	800 m
F.	Target Type	APC	Tank
G.	Masking	50-75%	None
H.	T/B Contrast	1.15	2.40
I.	Pattern Painting	Pattern	Solid
J.	Target Orientation	45 deg	90 deg
K.	Target Density	1 Target	3 Targets

The 2^{12-7}_{IV} experimental design was constructed from two 2^{11-7}_{III} basic and foldover blocks (Simon, 1973). This design was made up of 32 experimental conditions and was capable of estimating eleven main effects, fifteen strings of two-factor interactions, four strings of three-factor interactions (other than those confounded with main effects), and a block effect. Four measures taken on each experimental condition were combined into a proportion-of-targets-found score. No effort was made to minimize or control sequence effects. The experimental conditions and the performance scores are shown in Table 11.

Analyses of the 31 sources of variance are shown in Table 12.* In it are given the Effects, the Variance, and

---

*These are not the analyses found in the Navy report, which left much to be desired in this regard. The analyses and conclusions in this report are solely those of this author.

TABLE 11

EXPERIMENTAL CONDITIONS AND PERFORMANCE SCORES FOR NAVAL WEAPONS CENTER STUDY

Block I	(Basic 2_{III}^{11-7})		Block II	(Foldover 2_{III}^{11-7})	
1	ejk	.250	17	abcdfghi	1.000
2	afhi	.625	18	bcdegjk	.750
3	bfghk	.125	19	acdeij	.625
4	abegij	.750	20	cdfhk	.125
5	cfgij	.250	21	abdehk	.875
6	aceghk	.750	22	bdfij	0
7	bcehi	.250	23	adfgjk	.875
8	abcfjk	.625	24	deghi	.625
9	dghijk	.875	25	abcef	.750
10	adefg	1.000	26	bchij	.375
11	bdefik	.875	27	acghj	.500
12	abdhj	0	28	cefgik	.875
13	cdefhj	.625	29	abgik	.375
14	acdik	.250	30	befghi	.750
15	bcdg	0	31	aefhijk	.625
16	abcdefghijk	1.000	32	(1)	0

Order of Effects Across Design Matrix in Block I*

New Screening Label	A	B	C	D	E	F	G	H	I	J	K	Strings			
												(AD)	(BC)	(BD)	(CD)
Original Factorial Label	A	B	C	D	ABCD	ABC	BCD	ABD	ACD	AB	AC	AD	BC	BD	CD

*Block II is fold-over form. (See Simon, 1973)

TABLE 12. ANALYSES OF NAVAL WEAPONS CENTER EXPERIMENTAL DATA

Rank (largest 1st)	Source	Mean Difference (Effect)	Mean Square (Variance)	Eta Squared (η^2)	Cumulative Proportion of Variance Acct'd For
31	E	.3359	.9026	.2662	.2662
30	A	.2422	.4692	.1384	.4046
29	G	.2266	.4108	.1212	.5258
28	(AEF,...)*	- .2266	.4108	.1212	.6470
27	F	.1797	.2583	.0762	.7232
26	K	.1172	.1099	.0324	.7556
25	BC,DG,AF,HI,JK	.1172	.1099	.0324	.7880
24	D	.1016	.0826	.0244	.8124
23	FK,AJ,EI	- .1016	.0826	.0244	.8368
22	AK,EH,FJ	- .1016	.0826	.0244	.8612
21	BE,DK,GJ	.1016	.0826	.0244	.8856
20	AI,BG,CD,EJ,FH	- .1016	.0826	.0244	.9100
19	I	.0859	.0590	.0174	.9274
18	BK,CJ,DE	.0703	.0395	.0116	.9390
17	AE,HK,IJ	- .0703	.0395	.0116	.9506
16	Block Difference	- .0547	.0239	.0070	.9576
15	H	.0547	.0239	.0070	.9646
14	(ACE,...)*	- .0547	.0239	.0070	.9716
13	AB,CF,DH,GI	.0391	.0122	.0036	.9752
12	AC,BF,GH,DI	.0391	.0122	.0036	.9788
11	AH,BD,CG,EK,FJ	- .0391	.0122	.0036	.9824
10	B	- .0234	.0044	.0013	.9837
9	J	.0234	.0044	.0013	.9850
8	(ABE,...)*	- .0234	.0044	.0013	.9863
7	BJ,CK,EG	- .0234	.0044	.0013	.9876
6	EF,HJ,IK	.0234	.0044	.0013	.9889
5	CE,DJ,GK	- .0234	.0044	.0013	.9902
4	AD,BH,CI,FG	- .0234	.0044	.0013	.9915
3	C	.0078	.0005	.0001	.9916
2	(ADE,...)*	.0078	.0005	.0001	.9917
1	AG,BI,CH,DF	.0078	.0005	.0001	.9918

---

*Represents a string of three-factor interactions.

eta squared for each source along with the Cumulative Proportion of Variance Accounted For.

In the table, the sources have been ranked from the largest to the smallest effects. From that data, half-normal plots are supplied for this experiment (Figure 7-A) and for a second experiment (Figure 7-B) that was a repetition of the first but with different subjects. No other data is given here for the second experiment.

No detailed discussion of these results will be given here except to note that from an examination of all the data, it appears that at least four or five factors (E,A,G,F, and possibly K) out of the eleven appear to be critical. One three-factor interaction showed a large effect and a cursory examination revealed that out of the triple interactions in the string, one was Interaction AEF. Also the string of two-factor interactions showing the largest effect included Interaction AF. Since these both include the factors showing large (even larger) main effects, it suggests that both might be ordinal, and would not change the decision regarding the criticalness of any factor. From the half-normal plots (Figure 7), the only real difference between the results of experiments A and B is the increased importance of Factor K (target density).

No effort was made to discover why Factor K took on importance (i.e., eta squared = .139) for the second group, whether it was subject-by-factor interaction effect or the result of some unsystematic disturbance to the data. Factor D, on the other hand, was not considered a critical factor within the limits of the Experience (i.e., familiarity with the terrain) levels in this study, for three reasons: 1) it does not show up as a better-than-chance effect on either half-normal plot; 2) its effect is trivial in the second

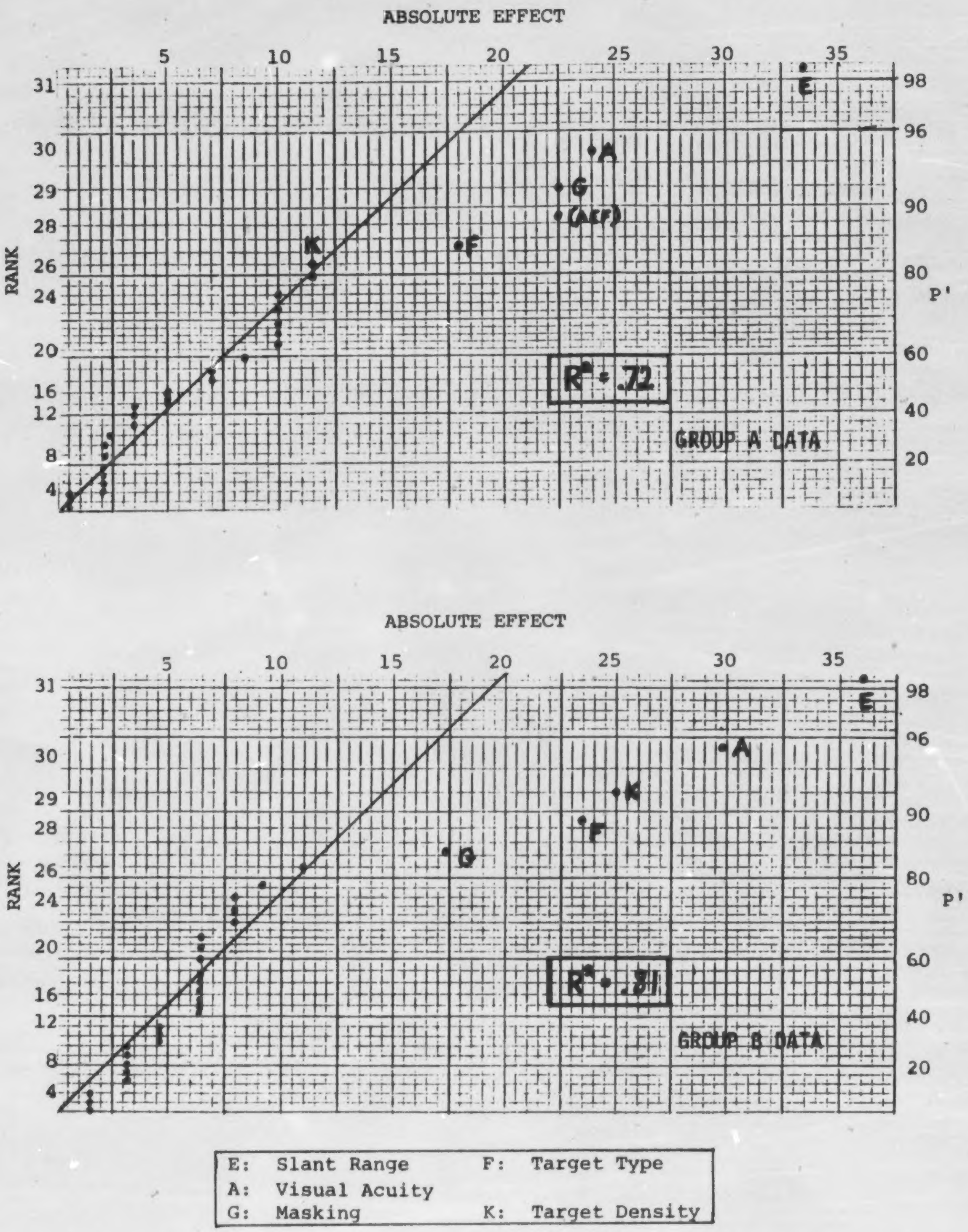


Figure 7. Half-normal Plots on Results of Naval Weapons Center Experiment.

experiment (i.e., eta squared = .002); and 3) in the design used for that experiment, the D effect could be severely confounded with a quadratic trend effect (i.e., 71%) if one exists. No center points were included in the original experimental design which might have provided a measure of trend through the data, as well as the basic for a test for lack of fit of the linear model of the screening design.

The investigators at the Naval Weapons Center ran another study using factors A,D,E, G* in a $2^2 \mathrm{x} 4^2$ factorial design and did an analysis of variance on the data. All factors but D were statistically significant at better than $p = .001$, while the F for Factor D was less than 1. The four factors plus several of their interactions accounted for .86 of the variance in that experiment, suggesting that the screening study was successfully picking important factors. Two-hundred fifty-six observations were required for this factorial study, and although functions were approximated through the mean data points for several pairs of factors, no overall function was calculated. Considerably more information in more useful form might have been obtained more cheaply had the original screening study been augmented with additional data points to create a central-composite design to be analyzed by a regression analysis.

---

*These letters refer to the factors as labeled in Table 10.

VI. ANALYZING SUBJECTS AS REPLICATIONS*

In the typical psychology experiment, when several subjects are tested on the same experimental condition,** the investigator will analyze the subject data by averaging each effect across subjects. Even when subject variance is isolated in these experiments, subject-by-factor interactions are usually included in the estimate of the "error" variance. This so-called error variance then is used to test the statistical significance of the estimated experimental effects. Of all the uses of subject replication, this most common use is probably the least informative.

When subjects are used in an experiment for replication purposes (which implies no interest in critical subject characteristics insofar as the replication group is concerned; the groups are presumed to be homogeneous), two kinds of analysis can be performed that will be considerably more informative than a test of statistical significance. In the early stages of the research program, the screening stage, where economy is being emphasized and little replication is anticipated, each subject as a replication who is added should represent a separate verification study. Each individual's data should be independently analyzed and the results compared among subjects. In this way differences due to a bad measure or to important subject-by-factor interactions can be detected rather than hidden among the averages. At the end of an experimental program, the data from subjects as replications would be used to establish confidence limits, which from an operational point of view is far more useful information than a test of statistical significance.

---

*This designation is used to distinguish this use of subjects from the case in which subjects are introduced into the experiment to represent specific combinations of subject characteristics. We expect subjects as factors to show a difference, or at least, would not be surprised if they do. On the other hand, we "hope" that subjects as replications will not differ in their performances, but would be interested to know if they do.

**In a survey of 239 experiments published in the Human Factors Journal, the Median number of subjects as replications was nine (Simon, 1976b, p 27).

ESTIMATING CONFIDENCE LIMITS

Cochran and Cox (1957, p 5) have this to say about significance tests and confidence limits: ". . . tests of significance are less frequently useful in experimental work than confidence limits. In many experiments it seems obvious that the different treatments must have produced some difference, however small, in effect. Thus the hypothesis that there is no difference is unrealistic: the real problem is to obtain estimates of the sizes of the differences. The construction of confidence limits may add something to the interpretation of a test of significance." They note that if the difference between performances on two machines is not found to be statistically significant, it does not prove that the performances (and thus the machines) are identical. They argue that if the 95% confidence limits for the differences in performance were relatively small, then the true differences would probably be of no practical significance, and ". . . consequently, it could be said that for all practical purposes the 2 machines are identical in speed. This is much more positive and useful than the mere statement that the difference in speeds was not statistically significant." Conversely, they add, if the confidence limits are large, then ". . . there is no justification for the conclusion that the machines can be regarded as equivalent. All that we have learned is that the data are not sufficiently accurate to show whether there is a difference in speed that is of practical importance."

In problems of equipment design, valid confidence limits are of considerable operational importance. While mean performance is useful to know, knowing the limits -- i.e., the estimated performance of the 95th or the 5th percentile man -- may be even more important from the standpoint of safety and/or mission success.

Confidence limits can be estimated with the following equation:

$$100\ (1 - a)\%\ \text{Confidence Limits} = \text{Mean} \pm \frac{t_a s}{n}$$

Where: t is the Student t for n-1 degrees of freedom at the error level

a is the probability of Type I error the investigator is willing to risk

n is the number of observations on the condition

s is the standard deviation of the replications

INTERPRETING MULTI-SUBJECT DATA

Subjects as replications should not be averaged together until it has been established that they are in fact homogeneous, at which time averaging becomes a cleaner way of handling the data although a less informative one. When subjects are used as replications, a complete analysis should be performed on the data from each one separately and the results compared. A number of possible outcomes may be anticipated, each with its own particular interpretation. For example:

1. The rank order of the different sources of variances (based on the magnitude of their effects) is essentially the same for all subjects.

2. A few sources are consistently ranked first for all subjects but after that there is little agreement.

3. The ranks agree among some subjects but not among others.

4. There is essentially no agreement in the ranks of the sources among subjects.

If the overall ranking of a majority of factors in a screening study agrees across subjects, there is reason for confidence that the results are probably accurate. It can be argued, of course, that just because two or three subjects agree that is no reason to believe that the results from 15 to 20 subjects would also agree. A sample of three, the argument goes, is just too small. It could, of course, be argued that in a population of thousands, 15 or 20 subjects are also a rather small number. However, it should not be forgotten that the purpose of this strategy is to check for gross errors and to do so as economically as possible. If, in fact, neither time nor economy are major considerations, then one might run the thousands of subjects. This still would not deny the importance of examining the results from each one at a time to find discrepancies. One strategy to increase one's confidence in the data from a few subjects is to select the few subjects at opposite extremes of skill or experience, for example, to test the limits. But when the agreement is good, for a screening study, only a few subjects (and a competent investigator) will ordinarily suffice.

As the differences in rank become more evident, more subjects may be required to understand why this is so.

If there is essentially no agreement in the source-ranks among subjects, it may be due to:

1. The collection of analysis of the data was sloppy with either considerable measurement or observation errors.

2. The performance measure may not be relevant to the problem or the task.

3. The factors actually have trivial or no effects on performance.

4. The task is either too difficult or too easy and little differentiation in performance is occurring.

When a few factors consistently rank first among subjects, but the remainder fail to agree, it is likely that those not agreeing are non-critical sources of variance and therefore show a variability both within and between subjects due to chance. The magnitude, as well as the ranks, will help determine if this interpretation is correct.

When the source-ranks among some subjects agree and disagree among others, several explanations are possible. For example:

1. If the results show several groups of subjects consistent within but not between groups, then it suggests that there may be unidentified subject factors interacting with the other factors. This is an important finding and should be investigated further.

2. If there is some consistency in the source-ranks among some subjects and no consistency observed among some others, this may mean that:

   a. The inconsistent subjects were doing so poorly that nothing really mattered.
   b. There were data-collection errors among the inconsistent.
   c. The inconsistent subjects had not stabilized their procedures before beginning the experiment and either changed their approach to the task in mid-study or exhibited learning (or fatigue) effects that distorted the experimental effects.
   d. The inconsistent subjects were tested across conditions in a different order and unisolated sequence effects might be distorting experimental effects.

Inspection of the raw data will often help find the explanation.

These are only a few possibilities. Only by inspecting the raw data before it is aggregated can an investigator begin to have faith in his results, particularly when the amount of data is small. Certainly when inconsistencies are observed, they should not be hidden by averaging on the assumption that this is a cleansing process. It is not. Averaging at the screening phase may hide important effects or the fact that the data is poor. Interpreting averaged results may lead to a distortion of the truth.

VII. ADJUSTING EXPERIMENTAL EFFECTS FOR TRENDS

Although the screening designs proposed in this report are robust to trends, when any overlap with a trend effect might distort the data more than is deemed incidental, the investigator may wish to adjust statistically the experimental effects for trends. An examination of the Percent Overlap data at the bottom of design matrices for the 16, 32, and 64 factors (Table 1, Appendices II and III, respectively) show which effects require adjustment. Even if the investigator has used procedures that are likely to minimize any trend effects, he may still wish to adjust as a precaution. It is apparent from the tables that those effects which must be adjusted for linear and cubic trends need not be adjusted for quadratic, and vice versa.

The methods of adjustment described here were taken from a paper by Daniel and Wilcoxin (1966)*. They applied the technique only to linear and quadratic trends. Methods for adjusting for cubic trends are also included in this report. When linear and cubic trends are both confounded with an effect, both must be adjusted simultaneously.

---

*Those who wish to refer to the original paper by Daniel and Wilcoxin (1966) to learn how the equations for the correction values are derived, will find the following pages in that paper the most informative. The general equation for deriving the correction factor for linear or quadratic trends is (4.10) on page 273; no equation was provided for calculating the cubic trend. The (L) term [or (Q) term] in that equation can be calculated from the sequence of identities (4.1) and (4.2) shown on page 269. It may also be calculated as the sum of the cross products between the particular integer Tchebycheff orthogonal polynomial coefficients and performance. Equations (4.7), (4.8) and so forth on page 272 are to be used to correct the appropriate estimated effects for trend. In Appendix VI of this report the derivations are given for the equations needed to adjust for both linear and a cubic trend together.

CONSTRUCTING AN ARTIFICIAL PROBLEM WITH TREND EFFECTS

To illustrate how corrections for trends are calculated and used, artificial data generated for a two-factor experiment, replicated twice, will be used. There are, therefore, eight observations and three effects, A, B, and AB. When the experimental conditions are arranged in the Standard Order, i.e., (1),a,b,ab,(1),a,b,ab, the performance scores, unbiased by trend effects, are:

$$-7, +1, -3, +9, -7, +1, -3, +9$$

respectively. These yield a mean performance of zero, regression coefficients of 5, 3, and 1 for the effects A, B, and AB, respectively, and no error. The equation formed from that data is:

$$Y = 5A + 3B + 1AB$$

To introduce trend effects into the data, linear, quadratic, and cubic coefficients of the integer Tchebycheff orthogonal polynomial (Fisher and Yates, 1963; Beyer, 1966; DeLury, 1950) were multiplied by a factor of -4, 2, and 1, respectively, and added to the experimental performance data. The total design, with supplemental data to illustrate how the performance data was produced, along with other calculations to be used later to adjust for trend effects, is shown in Table 13. The differences between the trend-free and trend-biased effects in this example are shown in Figure 8.

The trend-free and trend-biased performance values from Table 13 can be analyzed using Yates' algorithm to estimate the effects of A, B, and AB. These analyses are shown

TABLE 13. METHOD OF CONSTRUCTING ARTIFICIAL TREND-BIASED DATA

ORDERED EXPTL COND	TREND-FREE WEIGHTED EFFECTS IN FACTORIAL DESIGN									WEIGHTED TREND EFFECTS*				TREND-BIASED PERFORMANCE VALUES
	(I)	A	B	AB	(c	ac	bc	abc)	SUM_{ex}	L	Q	K	SUM_{tr}	$SUM_{tr} + SUM_{ex}$
(1)	0	-5	-3	+1	-	+	+	-	-7	-4(-7)	2(+7)	(-7)	+35	+28
a	0	+5	-3	-1	-	-	+	+	+1	-4(-5)	2(+1)	(+5)	+27	+28
b	0	-5	+3	-1	-	+	-	+	-3	-4(-3)	2(-3)	(+7)	+13	+10
ab	0	+5	+3	+1	-	-	-	-	+9	-4(-1)	2(-5)	(+3)	-3	+ 6
(1)	0	-5	-3	+1	+	-	-	+	-7	-4(+1)	2(-5)	(-3)	-17	-24
a	0	+5	-3	-1	+	+	-	-	+1	-4(+3)	2(-3)	(-7)	-25	-24
b	0	-5	+3	-1	+	-	+	-	-3	-4(+5)	2(+1)	(-5)	-23	-26
ab	0	+5	+3	+1	+	+	+	+	+9	-4(+7)	2(+7)	(+7)	- 7	+ 2
	LX	8	16	0	32	0	0	0		LL 168				LP -584
	QX	0	0	8	0	16	32	0		QQ	168			QP 344
	KX	16	24	0	-16	0	0	32		KK		264		KP 416
	Inner-products**									Sum of Squares**				Sum of Products**

---

[*] Numbers in parentheses are ordered integer Tchebycheff orthogonal polynomial coefficients for linear (L), quadratic (Q), and cubic (K) trends. Numbers in front of parentheses are the weights for each trend effect.

[**] Sum of squares, inner-products, and sum of products (e.g., LL, LX, or LQ) are derived using unweighted Tchebycheff coefficients.

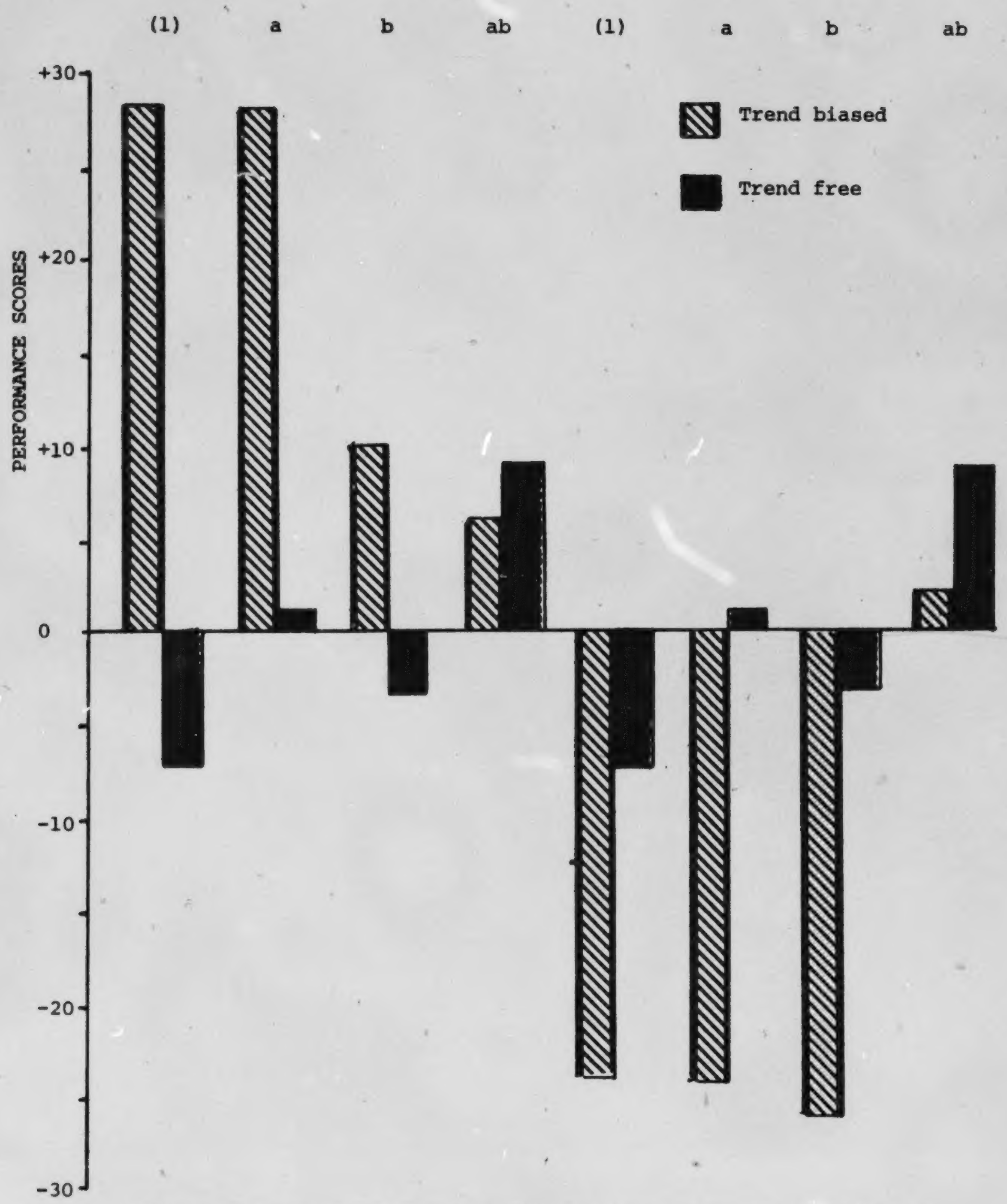


Figure 8. Comparing Results from Imaginary Experiment (Table 13), Both Trend-free and Trend-biased

in Table 14. A comparison of the coefficients of the three effects before and after they have been biased by trends reveals:

Effect	Trend-free	Trend-biased
A	5	3
B	3	2
AB	1	3

The differences are striking. In addition, no replication or replication interaction effects are indicated with the trend-free data but (as can be seen in Table 14-B) both show large effects in the trend-bias data. This would ordinarily be delegated to an error variance.

In practice, were this a real experiment, the investigator would have no idea what the true trend-free results should be. After all, the purpose of his experiment is to discover that from the sample data. All he knows are the performance scores and the results of their analysis. If he has no way of measuring the trend effects, or for that matter, even know for a fact that they exist, there is the real and ever-present danger that the only information he has will be distorted as in this example. Neither he nor his public, if he publishes, can know for sure. Eventually this distorted data becomes part of the lore naively referred to by some as "scientific" fact. It is not necessarily true, as some defenders of poor experimentation like to claim, that some information (however poor) is better than no information. When poor information can lead to erroneous decisions, it is better to have no information.

To offset these possibilities, the conscientious experimenter should first use procedures that help reduce or eliminate unwanted trend effects. Next, he should assign

TABLE 14

ESTIMATING EXPERIMENTAL EFFECTS IN THE ABSENCE AND PRESENCE OF TREND EFFECTS

Exptl. Cond.	Perform.	1	2	3 Effect-Totals	Unbiased Coeff. ÷8	Source
(1)	-7	-6	0	0	0	
a	+1	+6	0	40	5	A
b	-3	-6	20	24	3	B
ab	+9	+6	20	8	1	AB
(1)	-7	+8	12	0	0	Error
a	+1	+12	12	0	0	
b	-3	+8	4	0	0	
ab	+9	+12	4	0	0	

(14-A) EXAMPLE OF ANALYSIS OF DATA WITHOUT TREND BIAS

Exptl. Cond.	Perform.	1	2	3 Effect-Totals	Biased Coeff. ÷8	Source	Correction	Unbiased Coeff.
(1)	28	56	72	0	0			0
a	28	16	-72	24	3	A	- L-2C =	5
b	10	-48	-4	-16	-2	B	-2L-3C =	3
ab	6	-24	28	24	3	AB	- Q =	1
(1)	-24	0	-40	-144	-18	C	Repl (Error) =	0
a	-24	-4	24	32	4	AC		0
b	-26	0	-4	64	8	BC		0
ab	2	28	28	32	4	ABC		0

(14-B) EXAMPLE OF ANALYSIS OF DATA WITH TREND BIAS, SHOWING HOW BIAS IS CORRECTED his most important factors to trend-free columns. Then, as a final precaution, he should adjust the estimated effects for whatever trend remains. The method supplied by Daniel and Wilcoxin (1966) and supplemented by Webb (1977) is described next.

DETERMINING THE VALUE OF TREND-ADJUSTMENT FACTORS

If the investigator wishes to correct for linear or quadratic or cubic trend effects alone, each of these can be calculated independently of one another using equations I or II or III, respectively, in Table 15. If he wishes also to adjust for cubic trends as well, when they are correlated with linear trends, the adjustment factors for the two trends must be calculated simultaneously using the pair of equations, IV a and b in Table 15. The information required to solve these equations will be found in Appendices I, II and III as well as the equivalents of Table 14-B for new problems. In the discussion that follows, the artificial trend-biased data from the functional design described in the previous section will be adjusted for trends. How to perform the analysis when a screening design rather than a factorial design is involved will be discussed later.

Linear Adjustment Factor $\dot{L}$

Equation I in Table 15 is needed to calculate the linear adjustment factor. The numerical substitutions for symbolic values in this example are shown below:

$$[8(168) - (8)^2 - (16)^2]\dot{L} = 8(-584) - (8)(24) - (16)(-16)$$

$$(1024)\ \dot{L} = -4672 - 192 + 256 = -4608$$

$$\dot{L} = -4.5$$

TABLE 15

EQUATIONS NEEDED TO FIND THE LINEAR, QUADRATIC, AND LINEAR/CUBIC CORRECTION VALUES

I. Equation to Determine the Linear Correction Value $(\dot{L})$

$$[N(LL)-(LX)^2-(LY)^2-(LZ)^2\ldots]\,\dot{L} = N(LP) - (LX)(\underline{X}) - (LY)(\underline{Y}) - (LZ)(\underline{Z})\ldots$$

II. Equation to Determine the Quadratic Correction Value $(\dot{Q})$

$$[N(QQ)-(QX)^2-(QY)^2-(QZ)^2\ldots]\,\dot{Q} = N(QP) - (QX)(\underline{X}) - (QY)(\underline{Y}) - (QZ)(\underline{Z})\ldots$$

III. Equation to Determine the Cubic Correction Value $(\dot{K})$

$$[N(KK)-(KX)^2-(KY)^2-(KZ)^2\ldots]\,\dot{K} = N(KP) - (KX)(\underline{X}) - (KY)(\underline{Y}) - (KZ)(\underline{Z})\ldots$$

IV. Simultaneous Equations for Determining Linear $(\dot{L})$ plus Cubic $(\dot{K})$ Correction Values

a) $$[N(LL)-(LX)^2-(LY)^2-\ldots]\,\dot{L} - [(LX)(KX) + (LY)(KY) + \ldots]\,\dot{K} = N(LP) - (LX)(\underline{X}) - (LY)(\underline{Y}) - \ldots$$

b) $$-[(LX)(KX)+(LY)(KY)+\ldots]\,\dot{L} + [N(KK)-(KX)^2 - (KY)^2-\ldots]\,\dot{K} = N(KP) - (KX)(\underline{X}) - (KY)(\underline{Y}) - \ldots$$

SYMBOLOGY FOR TABLE 15

N	=	Total number of observations, $r2^{k-p}$, where p may be any value from 0 up to (k-1), and r is the number of times design is replicated.
L,Q, or K	=	Ordered integer Tchebycheff orthogonal polynomial coefficients for linear, quadratic, or cubic trends, respectively. (Found in Fisher & Yates, 1963; Beyer, 1966; DeLury, 1950).
LL,QQ, or KK	=	Sum of squared L, Q, or K Tchebycheff coefficients, respectively.
P	=	Performance values (as found in Table 14, second column).
LP,QP, or KP	=	Sum of cross products between Tchebycheff coefficients for a specific trend (L, Q, or K, respectively) and the corresponding performance values for the ordered experimental conditions.
X,Y,Z, etc	=	Ordered experimental conditions (± 1) for Effects X,Y,Z etc (as found in experimental design). (Number of effects involved depends on how many are correlated with particular trends being corrected.)
LX,LY,LZ, etc or QX,QY,QZ, etc or KX,KY,KZ, etc	=	Sum of cross products (called "inner products") between Tchebycheff coefficients for a specific trend (L,Q, or K, respectively) and the ordered experimental conditions (± 1) for Effects X,Y,Z etc (depending on how many are correlated with the particular trend being corrected). (Inner products for the designs in this report can be found in Appendix I-C, II-D, and III-D.)
$(\underline{X}),(\underline{Y}),(\underline{Z})$, etc	=	Effect-totals for Effects X,Y,Z etc (depending on how many are correlated with particular trend being corrected). (Effect-totals are found in the last column of Yates' analysis, before dividing by N, e.g. as illustrated in Table 14.)
$\dot{L}$, $\dot{Q}$, or $\dot{K}$	=	The unknown trend (L, Q, or K, respectively) correction value to be determined.

Note that while A, B, and AB are the only real experimental effects in this example, with eight observations, in theory, all effects of a 2^3 factorial can be estimated. For example, in Table 14-B we see that the effect-totals of the imaginary factor C is -144. In fact, this C represents a block effect, the difference between the two halves of the replicated experiment. At least one of the effects that is correlated with a linear trend cannot be used as an experimental factor in order to provide the necessary degree of freedom for the trend estimate. Factor C, the block effect, therefore would serve this purpose, it being the only remaining source of variance confounded with a linear trend.

Instead of using Equation I, Table 15, to make the calculation shown above, the adjustment for linear trend could have been done this way:

$$(1024)\ \dot{L} = LC(C)$$

$$(1024)\ \dot{L} = 32(-144) = -4608$$

$$\dot{L} = -4.5$$

In this calculation we used

$$[LC(C)] \text{ instead of } [N(LP) - (LA)(A) - (LB)(B)]$$

since they are equivalent. The equation on the right removes from the total, N(LP), the (LX)(X) terms of all sources of variance that were included as experimental factors correlated with linear trend (i.e., A and B). That would leave as a remainder, the value for all sources of variance that were not included in the experiment but were correlated with linear trend (i.e., C), which is what LC(C) represents.

Quadratic Adjustment Factor $\dot{Q}$

The calculations for isolating linear and quadratic adjustment factors are the same except that Q-values are substituted for L-values, as shown in Table 15, Equation II. The substitution of numerical for symbolic values in this problem are shown below:

$$[8(168) - (8)^2]\dot{Q} = 8(+344) - 8(+24)$$
$$(1280)\dot{Q} = 2752 - 192 = 2560$$
$$\dot{Q} = +2$$

As was done when estimating the linear trend adjustment factor, all the non-experiment sources of variance correlated with a quadratic trend could have been used to arrive at the same answer. For example:

$$1280\ \dot{Q} = 16(32) + 32(64) = 2560$$
$$\dot{Q} = +2$$

Cubic Adjustment Factor $\dot{K}$

This calculation would parallel the linear or the quadratic examples, except of course, only the terms that were a source of variance in the experiment and were correlated with a cubic trend would be involved. These are shown in Table 15, Equation III. The calculation would be:

$$[8(264) - (16)^2 - (32)^2]\dot{K} = 8(416) - 16(24) - 24(-16)$$
$$(832)\dot{K} = 3328$$
$$\dot{K} = 4$$

Linear Plus Cubic Adjustment Factors

Because the data was generated artificially, we know that both the linear and cubic adjustment factors just calculated are not correct. The linear one should not be -4.5 but -4, and the cubic should not be 4 but 1. These discrepancies occur because the linear and cubic trend effects are correlated with one another and if we intend to adjust the effects for both, then the adjustment factors must be determined for both simultaneously. This means that one may correct for linear and/or quadratic trend effects, but that if one were intending to correct for cubic and linear, the set of simultaneous equations, IV-A and B in Table 15 should be used to determine the pair of adjustment factors.* The substitutions of numerical for symbolic values in this problem are shown below:

a) $8(168) - (8)^2 - (16)^2 \dot{L} - 8(16) + 16(24) \dot{K} = 8(-584) - [8(24) + 16(-16)]$

b) $- 8(16) + 16(24) \dot{L} + 8(264) - (16)^2 - (24)^2 \dot{K} + 8(416) - [16(24) + 24(-16)]$

which can be simplified to:

$$1024 \dot{L} - 512 \dot{K} = -4608$$

$$-512 \dot{L} + 1280 \dot{K} = 3328$$

---

*These equations were derived by Dr. Steve Webb. The derivations are shown in Appendix VI.

If we multiply the second equation by two, we can eliminate $\dot{L}$, and solve for $\dot{K}$:

$$\begin{array}{rcr} 1024\,\dot{L} - 512\,\dot{K} & = & -4608 \\ -1024\,\dot{L} + 2560\,\dot{K} & = & 6656 \\ \hline 2048\,\dot{K} & = & 2048 \\ \dot{K} & = & 1 \end{array}$$

Substituting this in Equation IV-a,

$$1024\,\dot{L} - 512\,(1) = -4608$$

we simplify and get

$$1024\,\dot{L} = -4096$$

$$\dot{L} = -4$$

These values of $\dot{L}$,-4, and $\dot{K}$,1, are the weighting factors that we reused to create the artificial data.

Making the Adjustment for Trend

The equation for adjusting for any single trend effect is:

$$\dot{X} = \frac{(X) - TX\ (\dot{T})}{N}$$

where: $\dot{x}$ is the adjusted effect

(X) is the effect-total of the source being adjusted (e.g., A, B, AB, etc)

$\dot{T}$ is the particular trend correction factor (e.g., L or Q etc)

TX is the sum of the cross products between the coefficients of the particular trend and the source (e.g., LA or QA or LB, etc)

N is the number of independent observations in the experiment

For example, in our fictitious data in Table 14-B, the coefficient for the biased estimate of Interaction AB is 3. To correct that value for the bias introduced by the quadratic trend, we solve this equation:

$$\dot{AB} = \frac{(AB) - QX(\dot{Q})}{N}$$

$$\dot{AB} = \frac{24 - 8(2)}{8}$$

$$\dot{AB} = +1$$

The trend-free estimate of the coefficient for Interaction AB *is* 1.

To adjust an effect for *both* the linear and the cubic trend, the general equation is:

$$\dot{X} = \frac{(X) - LX(\dot{L}) - KX(\dot{K})}{N}$$

Thus to correct Factor A for both linear and cubic trend bias, we substitute:

$$\frac{24 - 8(-4) - 16\ (1)}{8} = +5$$

and for $\dot{B}$:

$$\frac{-16 - 16(-4) - 24(1)}{8} = +3$$

both of which are the coefficients we had derived before trends had been introduced to distort the data.

If one were to apply these same adjustments to the trend-biased effect-totals for the sources of variance associated with the block (replication) differences and each block-by-factor interaction, the corrected values would all be zero as they should be in our fictitious data.

Applying Trend-Adjustment Techniques to Screening Designs

Applying these techniques to screening designs involves no unique problems as long as the analysis is done with the original factorial labels in mind. The results from the Yates' algorithm will automatically rank the data in Standard Order using the original labels. These original labels are to be used as references to find in each screening design the values needed to make the trend adjustments for the particular effect. After the corrections have been made, the new screening design labels would be substituted for the original factorial labels.

VIII. HANDLING MULTIPLE RESPONSES IN SCREENING STUDIES

Human performance is situation-specific and complex. To understand and predict performance, therefore, it is necessary to examine all of the critical factors operating at the time performance is being measured (including those associated with antecedent events that also can affect performance). Equally important, but more frequently ignored, is the importance of providing measures that reflect the complexity of performance *in toto*.

Although methods of handling multiple performance criteria have been around for decades, experimental psychologists in general, and engineering psychologists specifically, have tended to examine the effects of experimental factors on multiple performance measures, a criterion at a time. As performance under operational conditions is generally complex, this one-at-a-time approach regarding responses is no more acceptable than it is regarding stimuli or the task situation. Informative results will be obtained only when it becomes common practice to perform *bilateral* *multivariate* experiments.

ADVANTAGES OF BILATERAL MULTIVARIATE EXPERIMENTS

The following are reasons why an investigator would want to include multiple responses as an integral part of his experimental plan and analysis:

1. A single measure usually does not adequately represent the typical complex performance under investigation.

2. A single measure may be an acceptable unitary concept but understanding would be improved if

it were broken down into component measures, rather than tying it to any single one.

3. Discovering evidence of interaction among response measures improves one's understanding of a phenomenon.

4. An analysis of multiple effects jointly may lead to different conclusions than would the sum of responses analyzed individually.

5. Understanding the joint contribution of several response variables can make it possible to select a smaller but most efficient combination of variables with which to measure performance.

6. It is more economical to carry out a single test rather than a number of separate tests for each response before a significant effect is detected.

7. Multiple responses increase the generality of the results.

8. A multiple response measure of performance in many situations is the more natural condition, whereas if efforts were made to hold some measures constant, artificial restrictions are introduced into the data to distort the interpretation. However, comparisons and assessments of factors and interactions when there are multiple responses are complicated by the fact that there is no unique linear ordering for vectors. Different approaches have been devised to overcome this.

The independent variables in screening designs are orthogonal (uncorrelated). However, it is highly likely that the dependent variables -- the responses, the criteria -- will be correlated to some degree.

Once an investigator has decided to make multiple responses a critical part of his investigation, he must then decide how he should analyze his data. It is not always obvious -- and in fact, it may be counterproductive -- to use the most sophisticated and formal methods of analysis available.

A statement by Gnanadesikan (1963, p 23) is appropriate here:

> *While the majority of multiresponse techniques, especially those in the formal framework of hypothesis testing, have been thought of as analogues of certain uniresponse procedures, yet from the standpoint of useful interpretations quite often these procedures are not such analogues . . . . It should, therefore, be emphasized that a multiresponse analysis should be considered as supplement to and not replacement for parallel uniresponse analyses. Methods which stimulate the user to look at subsets of responses, including the study of several responses individually, are thus very useful.*

In summary, the sophisticated investigator avoids a single cookbook analysis but instead examines his data with any technique that is likely to provide useful information.

Scope of This Section

A great many papers and books -- dating back to the mid-1930's -- have been written about the methods for analyzing experimental data involving multiple responses. In this section, therefore, no attempt will be made to explain the derivations of these methods in depth, nor to provide the reader with more than a cursory -- conceptual -- description of how to use them. The purpose of this section is to alert the experimental psychologist to the advantages of techniques of multivariate analysis and to encourage him to use them as a normal part of his experimental program. To do this, some of the more popular as well as some less familiar methods will be described. In some cases, enough information will be provided, hopefully, to take some of the mystery out of less familiar statistics, at least enough to make them easier to understand when the user must go to original papers to learn the mechanics of how to use them.

Some simple methods of analyzing multiple response methods are described because in many cases they will be more responsive to an investigator's needs than one of the more sophisticated analyses. For some of the more complex analyses, recent innovations that facilitate the interpretation of the data will be described. In some cases a method may be selected to avoid a large or unusual computer effort.

As Wilk and Gnanadesikan (1964, p 613) wrote:

> *. . . there is a long existant need for procedures to handle data involving multivariate responses in such a way that the resulting statistical summary and analysis (i) takes some account of the multivariate structure, and (ii) encourages insight*

> *into the experimental situation (as distinct from carrying out artificial and often pointless tests of hypotheses). The indefiniteness and complexity of objectives of statistical analysis of multi-response data emphasize the need for general informal procedures which help to convey to the data analyser some of the information implicit in the data.*

Hopefully, this section, while in many respects meager, will at least show the reader that there are choices to be made and provide enough detail to help him make the choice.

WEIGHTED CRITERIA

If the relative importance among n different sets of responses is known and can be quantified, the investigator can reduce the multiplicity of responses to a single value and treat the data as a unilateral analysis. For example, if all of the responses or criteria can be associated with a dollar value, or weighted according to their contribution to some other single concept, then they could be combined into a composite variate, W.

Before the weights are assigned, however, each set of performance scores must be transformed into standard scores. The standard score for each set of responses would be:

$$z_i = \frac{y_i - \bar{y}_i}{\delta_{y_i}} \quad \text{(i = 1 to n sets of responses)}$$

where $\bar{y}$ is the mean of the particular set of performance values. Each set of performance measures, Z_1 through Z_n would be assigned the weighted values b_1 through b_n

respectively and a composite score, W_i for each experimental condition would be calculated, thus:

$$W_i = b_1z_1 + b_2z_2 + \ldots.\ b_nz_n$$

GRAPHIC INSPECTION

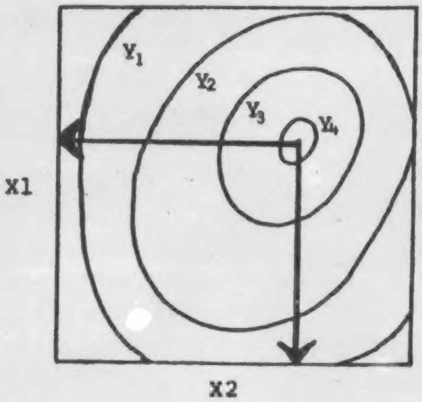
If the independent factors are quantitative and continuous, each performance measure may be presented graphically as a "response surface," which, with two predictor factors, has the appearance of a contour map with equal performance contours (e.g., Figure 9a). The hills and valleys of these response surface contours indicate the maxima and minima performance positions that can be associated with the coordinates (or values) of the independent factors. When optimum locations among multiple criteria do not coincide, the investigator must find a way of studying the data in order to make the best and most practical compromise.

If an investigator wished to find the optimum values of two predictor factors for a combination of performance measures, the contours for each measure could be drawn on a common coordinate system (e.g., Figure 9b). However, when there are more than two or three predictor factors, this graphic method becomes awkward to use unless it is meaningful to fix all but two of the predictor factors.

Given overlapping response surfaces, for example, one showing performance and the other showing costs, an investigator may visually search for the values of the equipment parameters (the predictor variables) that lead to some acceptable compromise between the two criteria.

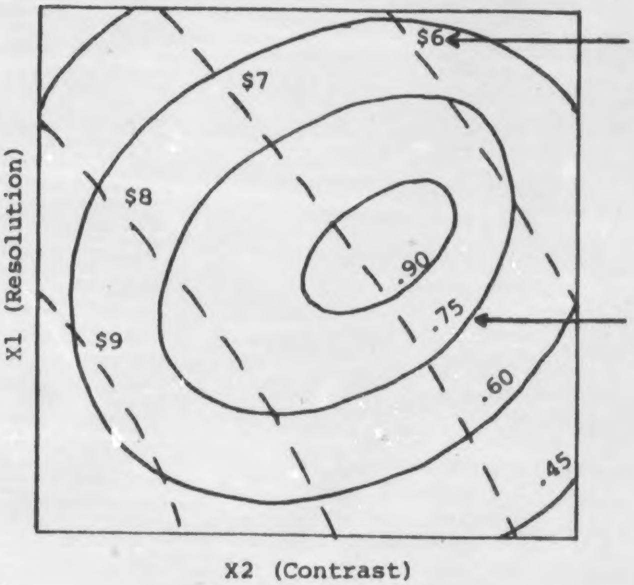
151

A



Arrows point to values of X1 and X2 that optimize Y

B



Dashed lines represent cost in thousands of dollars

Solid lines represent performance data

Figure 9. Artificial Data Illustrating Graphic Overlapping of Two Response Surfaces

USING LA GRANGE MULTIPLIERS

When there are too many independent variables to plot on a two-dimensional piece of paper (or attempt to draw as a three-dimensional surface), some technique other than overlapping plots of the response surfaces must be used. A procedure proposed by Umland and Smith (1959) may be employed. While their description treats the topic when only two criteria measures are being considered, it can be extended to handle more criteria.

They propose to use LaGrange multipliers* to find the optimum level of one fitted second (or first) order response function -- subject to the constraint provided by a second fitted second (or first) order response function. For example, assume we have two functions, one, the cost of building each particular equipment configuration (as represented by the experimental condition) and two, the level of operator performance at each condition. It would be possible to determine the combination of equipment parameters that optimized performance at some specified level while keeping the cost of the equipment within specified bounds. The converse could also be determined, i.e., the lowest cost for some fixed performance value. The procedure, a general outline of which is illustrated in Umland and Smith's (1959, pp 290-291) paper, is as follows:

---

*The general theory of LaGrange multipliers for solving constrained optimization problems is clearly presented in R. Courant, Differential and Integral Calculus, 1936, Vol. II, pp 188-202.

153

1. Two response surfaces are calculated using regression analysis to obtain the conventional least squares fit. Only first- or second-order surfaces can be handled, e.g., $\hat{Y} = \beta_0 X_0 + \beta X_i + \beta X_i X_j$
2. Differential equations are derived for each predictor factor in the two (or more) equations.
3. A new set of non-linear equations, using LaGrange multipliers is written.
4. These non-linear equations must be solved with one of a number of available computer programs. Umland and Smith (1959, p 291) suggest a method of steepest ascent as given by Booth (1955) for an IBM 650 Computer. However, a more recent program which Singer (1977) found useful was Subroutine ZXSSQ in the IMSL Library 1 (IBM 370 series computer).* Additional programming is required to fit the program to this particular application.

The results obtained would be the value of the two predictor factors for the optimum level of one criterion constrained by some value of the second.

Several precautions should be taken in using this technique:

1. An inspection of the surfaces individually will show whether they all have optima. Some surfaces appear as ridges rather than peaks which could cause the computer to either supply numerous correct answers or, more likely in the search mode, be unable to arrive at a solution.

---

*Institute of Mathematics and Statistics Libraries, Inc., Sixth Floor, GNB Bldg., 7500 Bellaire, Houston, Texas 77036.

2. Since optimum responses may not fall within the limits of the experimental space, limits must be written into the computer program to assure that the solutions obtained automatically by the computer will be useful.

3. Coding the independent variables can simplify the magnitude of certain calculations which may overload the computer.

STEP-DOWN PROCEDURE

If the investigator cannot assign quantitative values to his response, but is able to rank them in order of importance, he may assess the predictor factors in terms of the multiple responses as a series of single-response assessments, using a "step-down" procedure proposed by Roy (1958, p 1177) who notes:

> *The step-down procedure obviously is not invariant under a permutation of the variates and should be used only when the variates can be arranged on a priori grounds. Some advantages of the step-down procedure are (i) the procedure uses widely known statistics like the variance-ratio, (ii) the test is carried out in successive stages and if significance is established at a certain stage, one can stop at that stage and no further computations are needed, and (iii) it leads to simultaneous confidence-bounds on certain meaningful parametric functions.*

The investigator would use an ordinary F-test at each step of the analysis. He would begin by examining the most important response alone, and perform the analysis of variance and F-test on that. He would next use the second most important response to assess the data as a uniresponse analysis and F-test, but it would be conditional on the first response used. That is, he would perform an analysis of covariance, $y_{2.1}$, i.e., response y_2 with the effects of y_1 removed. Each succeeding response measure is made conditional on all previous response measures in the ordered sequence. This would continue until p response measures and p independent uniresponse assessments have been made.

Gnanadesikan (1963, p 23), in describing this technique, writes the following in regard to setting the probability value for rejecting the null hypothesis with this step-down procedure:

> The hypothesis for the multiresponse situation is not rejected if and only if none of the sequence of uniresponse hypotheses is rejected. Under the overall (i.e., complete multiresponse) hypothesis of no treatment effects, the separate F statistics are independently distributed. Hence, if $a_1, a_2, \ldots, a_p$ are the a-risks associated respectively with the p F-tests, then the overall a-risk is given by $1 - \prod_{i=1}^{p} (1 - a)$.

Roy (1958) describes how to choose the value of the a-risk (probability of error) at each step, so as to insure a desired overall a-risk for the combined data.

Gnanadesikan (1963, p 25) provides an example of this technique including a chi-squared-with-one-degree-of-freedom probability plot of the squared estimates of the different effects.

MULTIPLE ANALYSIS OF VARIANCE (MANOVA)

The use of MANOVA to analyze multiple response data is analogous to the use of analysis of variance to analyze single response data. The former takes into consideration the fact that multiple criteria are seldom completely independent and may depend upon one another or be hidden aliases of a single more fundamental criterion. As with ANOVA, an investigator may use MANOVA to:

1. Estimate the probability that two or more groups are really different, i.e., that an observed effect is a reliable one.

2. Determine the proportion of total variance accounted for by each factor, i.e., eta squared.

Instead of differences among means, we examine differences among centroids. Instead of studying the variance, we study the dispersion of the multiple responses in a multivariate space. Detailed discussions on MANOVA can be found in most references on multivariate analysis (e.g., Kerlinger and Pedhazur, 1973; Cattell, 1966; Cooley and Lohnes, 1971).

Making separate analyses for each of a number of response variables can lead to incorrect conclusions. Separate responses are seldom completely independent and in fact may be aliases of a single, more fundamental criterion. It is possible that no univariate criterion alone would distinguish among several groups, while a MANOVA would. This is illustrated with some fictitious data (Figure 10) taken from Kerlinger and Pedhazur (1973, p 359). It can be seen that when the means of conditions A_1, A_2, and A_3 are projected on either of the two dimensions, they are not well separated.

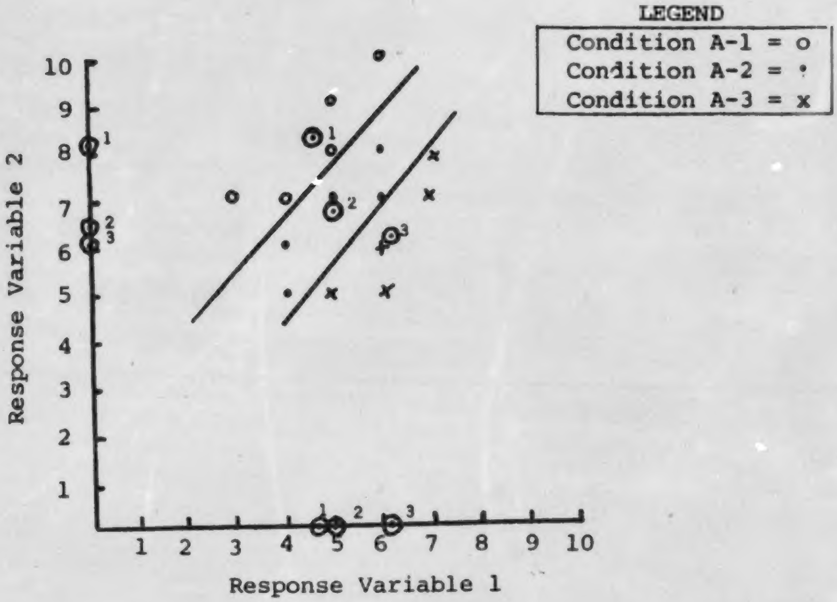


Figure 10. Illustration Showing How Analysis of Single Responses in Multiple-response Experiments May Fail to Detect Real Differences

[From Kerlinger and Pedhazur, 1973, p 359]

Yet an inspection of the two-dimensional plot shows that the three groups are clearly separated. This is what MANOVA would detect.

We will begin our discussion of MANOVA with eta squared, since in screening designs, this information would ordinarily be more important than significance tests.

MANOVA Eta Squared

Eta squared ($\dot{\eta}^2$) from multiple response analysis of variance problems will be calculated in one of two ways.

One-way designs. The first, which is not too important for screening studies, is used in a one-way design with only a single factor,

$$\dot{\eta}^2 = 1 - \frac{|W|}{|T|}$$

where $|W|$ and $|T|$ are the determinants of a within-treatment and a total-treatment matrix respectively.* This is analagous to the eta squared for the single-response ANOVA. Eta squared for ANOVA is equal to

$$\eta^2 = 1 - \frac{ss_w}{ss_t}$$

where ss_w and ss_t are within-group sum of squares and total sum of squares respectively. By subtracting that proportion of the total accounted for by the within group, i.e., $|W|/|T|$ from one, we have the proportion accounted for by the groups under consideration.

---

*In Appendix A of their book, Kerlinger and Pedhazur (1973) provide an easily understood short course in matrix algebra.

For a two-response study, the within-treatment matrix, W, and the total treatment matrix, T, would consist of the following elements:

$$W = \begin{pmatrix} ss_{w1} & sp_w \\ sp_w & ss_{w2} \end{pmatrix}$$

and the total treatment matrix, T,

$$T = \begin{pmatrix} ss_{t1} & sp_t \\ sp_t & ss_{t2} \end{pmatrix}$$

The elements of the matrices are calculated as follows. For a total of N observations, with r groups and n observations per group, the ss (sum of squares) and the sp (sum of products) are calculated in the conventional way. The total sum of squares would be:

$$ss_t = \sum_1^n X^2 - \frac{(\sum_1^N X)^2}{N}$$

Between-group sum of squares would be:

$$ss_b = \sum_1^r \frac{\sum_1^n X_i^{\ 2}}{n} - \frac{(\sum_1^n X)^2}{N}$$

Within-group sum of squares is obtained by subtraction:

$$ss_t - ss_b = ss_w$$

These must be calculated for each response measure (1 and 2, in our example which we call X and Y in our two-response example).

The sum of products (or sum of cross products between X and Y, or product-sum as it has been called) is calculated essentially the same as the sum of squares, except that instead of multiplying X times X to get X^2, we now multiply X times Y to get XY. Similarly, instead of multiplying ΣX times ΣX to get $(\Sigma X)^2$, we multiply ΣX times ΣY to get $(\Sigma X)(\Sigma Y)$. Thus, for a total of N observations, with r groups and n observations per group, the total sum of products would be:

$$sp_t = \sum_1^n XY - \frac{(\sum_1^n X)(\sum_1^n Y)}{N}$$

the between-group sum of products would be:

$$sp_b = \sum_1^r \frac{(\sum_1^n X)(\sum_1^n Y)}{n} - \frac{(\sum_1^n X)(\sum_1^n Y)}{N}$$

and the within-group sum of products would be obtained by subtraction, thus:

$$sp_t - sp_b = sp_w$$

Within each matrix, the sums of products in corresponding positions on either side of the main diagonal, are the same (since sp_{12} is the same as sp_{21}).

Multifactor (and screening) designs. When the study involves more than one factor, as in the screening design, and there are multiple responses, the equation for eta squared is

$$\eta^2 = \frac{|F + E|}{|T|}$$

Note once again the analogy between this eta squared for multiple-response data and eta squared calculated from single-response data. For single response data, eta squared would be the ratio of the sum of squares for the particular factor or interaction over total sum of squares. In the multiple-response case, it is the ratio of the determinants of the factor matrix (F) plus error (E) matrix over the total matrix.*

For screening designs, the F-matrix represents both main and interaction effects. The E-matrix is equivalent to the W-matrix in the previous equation for eta squared, both being the residuals after all sources of variability between groups have been removed from the total variance, or dispersion. Thus, in MANOVA with multifactors, the between groups dispersion can be partitioned into matrices for the individual factors and the interactions, and eta squared values determined for each of them as in the ANOVA case. Of course, with Resolution III designs, interaction terms are not isolated from main effects. With Resolution IV designs the two-factor interactions are in fact strings. This does not change the calculations. For MANOVA, the sources of variance are partitioned in the same manner as in single response ANOVAs. In a two-response study, for example, the matrix for Factor A would look like this:

---

*It should be noted that in the first equation it is necessary to work from the within matrix rather than get eta squared from a between matrix directly. In the second equation, it is necessary to add the error matrix to the particular factor matrix before finding the determinant. These are necessary because all between-treatment matrices (which include a factor or interaction matrix) are singular. That means that at least two columns (or rows) of the matrix are proportional to one another, e.g., 1 2 3 and 2 4 6; the determinant of a singular matrix is always zero. This "no solution" situation is avoided by working with the within -groups and then subtracting, or by adding the error matrix to the between-matrix. Because of this restriction, no eta squared can be calculated for a screening design unless it is repeated at least twice and an error term is obtained. At least, the author was unable to find another solution by the time this report went to press.

$$A = \begin{pmatrix} ss_{A1} & sp_{A} \\ sp_{A} & ss_{A2} \end{pmatrix}$$

and for Interaction AB, for example, like this:

$$AB = \begin{pmatrix} ss_{AB1} & sp_{AB} \\ sp_{AB} & ss_{AB2} \end{pmatrix}$$

Elements in the matrices for main effects are calculated in the same manner they would be for the between-treatments matrix. The only new elements are those for the interactions, and these are not difficult to calculate with screening designs in which all the interactions are linear products of two two-level main effects. Thus the same equation is used to calculate each element of the interaction matrices as the main effects. The only difference in the calculation is that with main effects, the ΣX_i and ΣY_i , represent the summing of performance scores obtained under all high or all low conditions, while with interaction effects one would sum either all conditions in which both factors levels were high and both were low, or one would sum all conditions in which the factors levels were always mixed, one high and one low.

These two sums now represent the sum of two "groups" from which sums of squares and sums of products are calculated.

Once the appropriate sum of squares and sum of products are obtained, the equation for eta squared requires that matrices F and E be added. To add two matrices, in this case F and E, it only is necessary to add the elements in corresponding positions in each matrix to form the matrix sum. For example:

$$\begin{pmatrix} 3 & 5 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 8 \\ 5 & 9 \end{pmatrix} = \begin{pmatrix} 4 & 13 \\ 7 & 15 \end{pmatrix}$$

with 4 obtained by adding 3 plus 1, and 13 obtained by adding 5 plus 8, and so forth. (You can not obtain the determinants for F and E and add them to get the determinant for the sum. One must sum first and then get the determinant.)

In Appendix VII, algebraic equations are given to calculate the determinants for 2 x 2 and 3 x 3 matrices, used when there are two or three responses in the MANOVA. When there are more responses, the analysis is sufficiently complex to require a computer.

Multi-variate Test of Significance

In multivariate analyses, much attention -- possibly too much attention -- has been directed at tests of statistical significance. Such tests, for a null hypothesis of "no effect" against the completely general alternate hypothesis, have important limitations. While a number of tests have been devised, choice among them is based largely on intuition.

Wilks' lambda (Λ) (generalized mean) test is one of the more popular tests of significant differences between groups in multiple response studies and will be described here. It determines a probability level for the null hypothesis of equality of population centroids (mean vectors) on the assumption of equality of dispersion (variance-covariance matrices). The assumption is analogous to that of homogeneity of variance in the univariate F-ratio test of equality of means.

The equation for Wilks' lambda is:

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|B + E|} = \frac{|E|}{|F + E|}$$

Matrix T is equal to matrix (B + E), which is not surprising since the total is equal to the between plus the within. We have already indicated that both W and E are residual matrices that are left after all known sources of variance have been removed from the total.* In multifactor designs the B-matrix would become a matrix (F) for each particular factor or interaction.

Although the explicit distribution of Wilks' lambda is not known except for a few special cases, there are a number of transformations which enable lambda to approximate the classical F-distribution. Most of them, as given, are usually suitable only for the one-way MANOVA design. Tatsuoka (1971, p 200) gives the formula for Rao's R-statistic having an approximate F-distribution which is suitable for the multiple independent variable (and screening design) case, provided there is an estimate of error variance and co-variance possible. The equation he gives is:

---

*We shall assume that we are always dealing in screening designs with a Model I (final effects) experiment.

$$R = \left(\frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}}\right)\left(\frac{ms - (pv_h / 2) + 1}{pv_h}\right)$$

with $m = v_e + v_h - (p + v_h + 1)/2$

and $s = \sqrt{\frac{(pv_h)^2 - 4}{p^2 + v_h^2 - 5}}$

with pv_h and $ms - (pv_h/2) + 1$ degrees of freedom. Also

v_e = Number of observations in basic screening design multiplied by number of repeats beyond the original plan.

v_h = Number of groups in factor being investigated, minus one. In screening designs this value will be 1 for main and interaction strings.

p = Number of dependent variables.

MANOVA Versus Multiple Discriminant Analysis

Although it is not the intention in this report to review every form of multivariate analysis available, some comments regarding multiple discriminant analysis as it relates to MANOVA may be helpful. Both techniques may be used to examine one-way designs (single factor, multiple conditions) with multiple response data. For a given set of data, both techniques will produce identical overall tests of statistical significance.

But MANOVA stops with this test of significance, while multiple discriminant analysis provides the user with some indication as to the nature of the difference. It does this by providing a set of weights or coefficients for the several dependent measures that will separate the mean values of the conditions to the maximum extent. Essentially what is happening is that they are turning the original dependent variables into new orthogonal dimensions (i.e., canonical variables) which, like the factors of factor analysis, may not be readily named. In certain human factors for equipment design problems one may not find the orthogonal, artificial variables as useful as the real world ones. The canonical variables may provide clues for better understanding, yet the original variables may still be of greater practical value. Multiple discriminant analyses were developed to handle one-way designs. In a multiple-response, multifactor screening design, there are separate discriminant analyses, one for each main and interaction effect. Multiple discriminant analysis can be found in most books on multivariate techniques (e.g., Cooley and Lohnes, 1971; Kerlinger and Pedhazur, 1973).

GRAPHICAL ANALYSIS USING ORDERED DISTANCES

Wilk and Gnanadesikan (1961; 1964) describe a procedure for graphical analysis of multiple response data by means of "probability plots." Their procedure represents a generalization and an extension of the technique of half-normal plotting proposed by Daniel (1959) for the graphical analysis of single-response data. It was proposed specifically to be used with two-level factorials where there is a meaningful decomposition of the treatment structure into orthogonal single degrees of freedom contrasts. It can also be applied to results from the fractional factorial and screening experiments. Where no independent estimate of error is available, the use of this "internal comparison"* method has several advantages:

1. It may reveal significant effects when single-response analysis does not.
2. It may lead to smoother, more stable statistical configurations than a single-response analysis.
3. It provides an easily assimilable summary of experimental results that facilitates investigator personal inspection of the data.

---

*"Internal comparison" refers to comparisons based on a statistical standard set by the data.

4. It helps clarify the interpretation of data when different responses are not orthogonal to one another.

Throughout the many references to this technique, the point is made continually that the intent is not to supplant the marginal analysis of individual responses. Instead, both types of analysis should be used to supplement one another. Roy, Gnanadesikan, and Srivastava (1971, pp 97-112) devote an entire chapter to graphical methods and internal comparison evaluation procedures for multiple response data, including examples.

General Description

Analogous to the case of the half-normal plot, the multiple response method of graphical analysis is based on probability plots of ordered squared distances (defined as "positive semi-definite quadratic forms"). Ordered distances are judged to be real when they deviate considerably from a straight line plotted on appropriately scaled paper. Several problems arise, however, with multiple response analysis that are not present in single response analysis. One, in multiple response analysis, it is necessary to approximate and estimate the distribution which serves as the appropriate basis for the probability plots. A procedure for doing this may be based on order statistics from the gamma distribution and tables to facilitate the required estimation. Two, while the univariate analysis may be based on the half-normal distribution (i.e., chi-square distribution with one degree of freedom), the multivariate analysis uses the standardized gamma distribution of a particular shape determined by the data. Three, unlike the univariate case, the problem of linearly ordering multivariate data is complicated by the lack of a convenient measure of "size." Gnanadesikan and his co-workers have developed techniques to help solve these problems. Only a general description of these techniques will be supplied here. The reader is referred to the original papers and other references on the topic for a working knowledge.

Gamma distribution paper. This technique requires that the squared distances be ordered and plotted against the corresponding quantiles of the gamma distribution. Psychologists are familiar with special cases of the gamma distribution, e.g., the chi-square and exponential distributions. Unfortunately, unlike the uniresponse procedure proposed by Daniel for which special "probability" paper can be prepared, no single general probability paper can be prepared for the gamma distribution. This is because the distribution can be standardized through a linear transformation for only two of the three parameters defining the distribution, that is, for the origin and the scale, but not for the shape. Special approximation tables or a high-speed computer are required to calculate the actual percentage points of ordered effects. Wilk, Gnanadesikan, and Huyett (1962) and Roy, Gnanadesikan, and Srivastava (1971) provide tables of percentage points for the reduced gamma distribution, together with the numerical procedures and approximations employed. Wilk, et al (1962, pp 102-103) describe the procedure step by step and note that the entire procedure is mechanized and in use at Bell Telephone Laboratories for the IBM 7094 and GE 635 computers. Computer programs for these calculations are also given in Roy, et al (1971).

Calculating the Ordered Distances

The effect of a factor in the univariate case is the mean difference in performances between high and low levels of the factor. With multiple responses, the measure of the main effect of a factor would be the "distance" between the high and low centroids in the multi-dimensional response surface. For example, if there were three independent factors with two levels in each and two responses, one might graphically represent the data as shown in Figure 11. The performances on conditions involving high and low levels for Factor A are indicated by squares and circles, respectively. The centroids are the darkened symbols. Roy, et al (1971) describe the calculation this way:

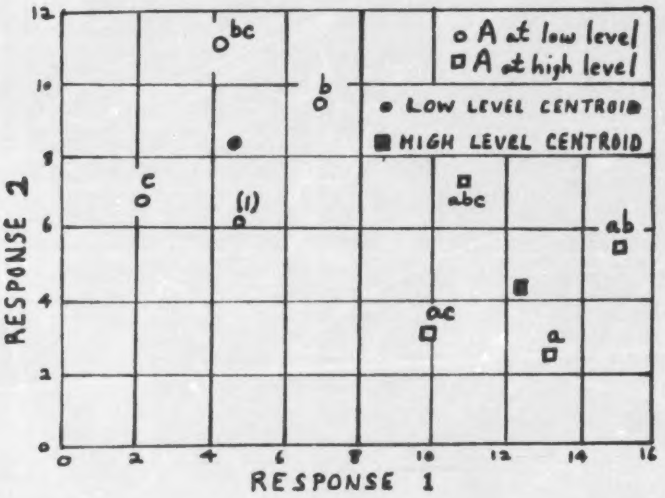


Figure 11. Geometric Representation of One Main Effect, A, in a 2^5 Experiment with Two Responses. [From Wilk and Gnanadesikan (1964, p 619, Fig. 1.]

For the bivariate response, therefore, a natural measure of the main effect A would be the "distance" between the centroids in the two-dimensional response space. If x_1 is the contrast vector corresponding to the main effect A, then the "distance" between the two centroids is proportional to the "length" of x_1. For instance, choosing the compounding matrix A, in the defining equation

$$\left[d_i = x_i A x_i', \quad i = 1,2,\ldots L(\leq n-1) \text{ responses} \right]$$

as the identity matrix of order 2 in this case, so that $d_i = x_1'x_1$, we get the squared Euclidian distance between the two centroids corresponding to the definition of the main effect A. More generally, the (n-1) contrast vectors x_1's may be visualized as (n-1) points in the p-dimensional space, as squared lengths, or squared distances from the origin, associated with the contrast vectors.

Selecting the compounding matrix. The defining equation, written with matrix symbols, can be expanded to look like this:

$$\text{Squared Distance, } d_i = \begin{bmatrix} x_1, x_2, \ldots x_L \end{bmatrix} \underbrace{\begin{bmatrix} a_1 & & a_{ij} \\ & a_2 & \\ & & \ddots \\ a_{ij} & & a_L \end{bmatrix}}_{\boxed{\text{COMPOUNDING MATRIX}}} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_L \end{bmatrix}$$

It is necessary for the investigator to arbitrarily specify the values of the a weights of the compounding matrix with the single restriction that the squared distances are greater than or equal to zero. Symbolically:

$$x'Ax \geq 0$$

Wilk and Gnanadesikan (1961, p 1210) state that the elements in the A matrix are non-negative definite quadratic forms. Some possible examples of the A matrix might be a) the identity matrix, (I), b) a diagonal matrix of reciprocals of estimates of the variances of the p responses, ($D\frac{1}{s_{ii}}$), or c) the inverse of the covariance matrix of the original responses, (S^{-1}).

The inverse of the covariance matrix, S^{-1}, is a particularly useful compounding matrix since it provides a linear invariance and makes statistical allowance for differing variances and correlations among the elements of the effects vectors. However, it is recommended that an S^{-1} matrix be derived from the sum of squares and sum of products of r effects (contrast) vectors, where r is a subset of the total number of effect vectors. In the case of ordered values, the subset of r vectors might include the smaller half of the effects. This removes the larger effects from the estimates, for if they are real, including them would reduce the number of effects that would appear to stand out from the rest. Excluding them gives the smaller, but real effects a better chance of being detected.*

Two other useful compounding matrices, I and $D\frac{1}{s_{ii}}$, are diagonal matrices. The diagonal matrix with weights inversely proportional to estimated variances, has been found to yield a more sensitive analysis than equal weighting as long as the estimated variances are based on the smaller half of the ordered effects vectors (as proposed for S^{-1}).

Roy, et al, recommend that several different compounding matrices be tried in estimating the squared distance and the researcher should realize that whatever compounding matrix is used, subsequent inferences regarding the data should be "conditional" on this choice.

---

*Note the similarity between that tactic and that proposed by Zahn when he calculates the standard deviation for the half-normal plot (see p 97, this report).

Analyzing Subgroups

Graphical internal comparison procedures may also be applied to subgroups of the effects vectors, selected according to meaningful criteria which are independent of the data. For example, one might look at different orders of effects separately, e.g., main and two-factor interactions, or isolate all higher-than-second-order interactions and examine them.

Plotting and Evaluating the Ordered Distances

It has already been stated that under the null hypothesis, i.e., no systematic effects, the ordered distances would behave like a random sample from a gamma distribution with its density defined by origin, scale, and shape parameters. By keeping the origin at 0 and the scale at 1, only the shape parameter is unknown. If it were known, then when the ordered distances were plotted against corresponding quantiles of the gamma distribution, the points would appear in a straight-line configuration if there are no real effects. Major departures from the straight line by the largest effects will suggest that those effects are probably real.

Conclusion

While there is much to learn before one can comfortably use this graphical, internal comparison method, there seems to be sufficient justification to apply it to screening problems. Since without replication, the screening plans have no independent estimate of error variance to test the significance (reliability) of an effect, this internal comparison procedure serves as a useful alternative. Before anyone can assess how valuable the technique is, more experience is needed in using it and applying it to behavioral data.

CANONICAL CORRELATION ANALYSIS

Canonical correlation analysis is the generalization of univariate multiple correlation analysis to two sets of variables, usually, but not always, multiple independent and multiple dependent variables. Canonical analysis provides a measure of the degree of association between the two sets of variables and may be useful for learning something about the underlying relationships among the variables of the two sets.

Applications

Examples of two sets of multivariate data to which canonical correlation analysis might be applied to determine the degree of association and underlying relationships are:

1. Flight performance measures at the beginning and the end of a training program.
2. Instructors' characteristics versus trainees' flight performance measures.
3. Instrument design factors versus multiple cost criteria (e.g., dollars, performance).
4. Pilot selection test scores versus flight performance data.
5. Pilot training-simulator design parameters versus multiple transfer-of-training criteria.

Process

Many books have been written on canonical correlation analysis, the theory and the mathematics, (e.g., Cattel, 1966; Kerlinger and Pedhazur, 1973; Nie, Hull, et al, 1975; Bock and Haggard, 1968; Tatsuoka, 1971). These will not be discussed here. Although that background is important, at the end of this section an improved canonical analysis will be described. Therefore at this time, only the fundamental process involved in the canonical analysis will be discussed.

We begin with a table showing the coordinates of the experimental space at which the data was collected and the set of measures made on that set of conditions. In screening designs, the coordinates are the conditions of a fractional factorial and therefore, orthogonal. The response measures are almost always correlated. Thus the raw data matrix for three independent and two dependent variables would look like this:

Observation Number	Set 1 Independent			Set 2 Dependent	
	A	B	C	X	Y
1	-1	-1	-1	.3	14
2	+1	-1	-1	.7	21
3	-1	+1	-1	.1	13
4	-1	+1	-1	.5	11

...etc etc etc...

From this data a table of intercorrelations is constructed by finding the correlation between every pair of columns, and locating them in the intercorrelation table as follows:

$$\begin{array}{cc|ccc|cc}
 & & \multicolumn{3}{c|}{\text{Independent}} & \multicolumn{2}{c}{\text{Dependent}} \\
 & & A & B & C & X & Y \\ \hline
 & A & r_{aa} & r_{ab} & r_{ac} & r_{ax} & r_{ay} \\
(I) & B & r_{ba} & r_{bb} & r_{bc} & r_{bx} & r_{by} \\
 & C & r_{ca} & r_{cb} & r_{cc} & r_{cx} & r_{cy} \\ \hline
(D) & X & r_{xa} & r_{xb} & r_{xc} & r_{xx} & r_{xy} \\
 & Y & r_{ya} & r_{yb} & r_{yc} & r_{yx} & r_{yy}
\end{array}$$

which can be simplified using matrix algebra and symbols as:

$$R = \left[\begin{array}{c:c} R_{11} & R_{12} \\ \hdashline R_{21} & R_{22} \end{array}\right]$$

where R is the entire correlation matrix, R_{11} represents the correlations among the independent variables, R_{22} represents the correlations among the dependent variables, R_{12} represents the correlations between independent and dependent variables, and R_{21} represents the transpose of R_{12}.

Computer programs exist that would work from the data in the above matrix to find the solution to the canonical correlation analysis. This in essence is what it would do. It would search out a set of weights (i.e., Beta coefficients) to assign to the independent variables and another set of weights to assign to the dependent variables. With these, two sets of canonical variates would be calculated. A "variate" is a rotated dimension in the multivariate space made up of composite scores derived from the weighted values of the two sets of raw data.

The weights for the two sets are selected in a way that will cause the correlation between the pair of variates to be a maximum. The square of this correlation indicates the proportion of the variance of the single criterion composite accounted for by the predictor composite.* Next, a second pair of variates could then be calculated that would account for as much as possible of the variance between the two sets that were left unaccounted for by the first pair of variates. This procedure can continue, the maximum number of iterations being equal to the number of variables in the smaller of the two groups. Each new pair of variates is completely orthogonal to all previous pairs of variates. It may not be necessary to complete them all since most of the variance may be accounted for by the first few pairs.

Since the new variates are formed in pairs, the existence of large weighting (coefficients) on the old variables in the two groups would identify which ones were responsible for the degree of correlation that was found. For example, an idealized result might be:

	Old Variables	New Variates	
		I	II
Group I	1	H	L
	2	H	L
	3	L	H
	4	L	H
Group II	5	L	H
	6	L	H
	7	H	L
	8	H	L
Canonical Correlation:		(.85)	(.75)

Coefficients

H = high weight

L = low weight

---

*Thorndike (1975) discusses general considerations in interpreting canonical correlations and specifically (pp 82-83) some problems in interpreting the index of proportion of variance. A "redundancy index" is proposed instead.

These results could be interpreted as follows. In the first variate (I), approximately 72% of the variance due to Variables 1 and 2 were accounted for by Variables 7 and 8. In the second variate (II), approximately 56% of the remaining variance (after Variate I was discounted) in Variables 3 and 4 were accounted for by Variables 5 and 6. As in factor analysis, it may be possible to find the common element among the heavily weighted variables to be able to name the variates in the two groups of data.

Limitations of Canonical Correlation Analysis

With real data, these clear cut divisions and associations found in the above example seldom occur. The problem of interpretation may be difficult. Trying to "name" the new variates may also be difficult.

Perhaps the major limitations of a canonical correlation analysis lies in the unreliability of the weights. The problems that arise in trying to examine the coefficients of individual terms in multiple regression problems when the variables are correlated are only complicated further in these bilateral regression analyses. Hoerl and Kennard (1970a, b), cite the following characteristics of coefficients estimated from ill-conditioned experimental designs:

1. The coefficients become too large in absolute value.

2. Some coefficients have the wrong sign.

3. Collectively the coefficients are unstable; another set of performance data would be unlikely to give the same beta values.

4. Individual coefficients may be over or under estimates of the strength of a particular factor.

To try to interpret the results from a canonical correlation analysis by examining the individual weights, therefore, seems to be overly optimistic. The more non-orthogonal the original matrices, the less reliance can be placed on the interpretation of individual coefficients. (See Simon, 1975 for more discussion of this problem.)

An Improved Method of Canonical Correlation Analysis

Hoerl and Kennard (1970a, b) proposed to use "ridge regression" to improve the analysis of an ill-conditioned multiple regression matrix. This analysis, they suggest, will obtain a better prediction equation in which:

1. The estimated coefficients will be closer to the true coefficients on the average.

2. The signs attached to the coefficients will be more meaningful.

3. A point estimate of a response can be made with a smaller mean square error.

4. The coefficients will be more stable and likely to be repeated if new data is taken.

Hoerl and Kennard's (1970a, b) original papers provide a description of the philosophy and underlying mathematics for ridge regression analysis. A simpler explanation has been provided by Simon (1975) and will not be repeated here. Mechanically what is done is to add a small constant to the unit diagonal of the intercorrelation tables, and then analyze this modified data by a multiple regression analysis as usual. Finding the proper constant (usually less than .05) depends on a study of a plot of the coefficients obtained with each constant after trying a range of values. A number of studies have found that for highly correlated matrices, ridge regression analysis provides a more stable set of coefficients and a smaller prediction error than conventional multiple regression.

Carney (1975) proposes using ridge regression analysis rather than multiple regression analysis to obtain canonical correlations. As with the single response case, this would reduce the instability and the errors in the estimates of the weights used to obtain the canonical variates. He developed a computer program that would provide Monte Carlo data to evaluate and solve the "canonical ridge estimates" (Carney and Anderson, 1974).

The investigator must decide what constant, k, to add to the unit diagonals of two matrices, R_{11} and R_{22}, for the canonical ridge analysis. Carney (1975, p 9) says: "There seems to be no theoretical criterion for choosing k-values for canonical ridge estimates" but he suggests several possible empirical approaches:

1. Try a series of k-values and select the solutions in which the coefficients appear not to change much over a range of k's. (This is feasible for ridge regression, with a single set of coefficients, but can be more difficult with the many coefficients in the canonical ridge case).

2. Limit the application of ridge to the first canonical correlates only.

3. Proceed as in the Monte Carlo experiments, treating the sample covariance matrix as if it were a population matrix, generating artificial samples, and selecting k-values to minimize "mean square error."

4. Perturb the data matrix and attempt to find k-values for which the perturbations have little effect.

5. Subdivide the sample and select k-values for which stability across subsamples occur.

IX. EVALUATING THE ADEQUACY OF THE REGRESSION EQUATION

One of the better features of central-composite designs is the procedure that enables the investigator to:

- Collect data sequentially in blocks, beginning with only enough for a first-order model when no function is assumed
- Determine whether the order model adequately fits the actual data
- Collect more data when lack of fit is significant in order to fit the next higher-order model.

The analysis of variance of the classical central-composite designs (Box and Wilson, 1951; Box and Hunter, 1956; Simon, 1970b, 1973), composed of 2^{k-p} fractional factorials and center points in the first-order model plus "star" points in the second-order model, would ordinarily take the form of these examples:

First Order (3 factors, 4 center points, 12 observations)

Source	d.f.	
First order terms		3
X_1	1	
X_2	1	
X_3	1	
Lack of fit		5
Error		3

---

*Most of the material for this section was taken from a paper by Draper and Herzberg (1971). Mr. Edward J. Dragavon helped interpret the paper and prepare the example.

Second Order (3 factors, 6 center points, 30 observations)

Source	d.f.
First order terms	3
Second order terms	6
Lack of Fit	5
Error	5

Draper and Herzberg (1971) show how the lack of fit in each of these two types of designs -- first or second-order -- can be split into two sources that can help the investigator decide where the lack of fit (bias) lies and what his next step should be.

SPLITTING THE LACK OF FIT OF THE FIRST-ORDER DESIGNS

The sum of squares for the first-order lack of fit can be split into:

L_1: Sum of squares due to lack of fit of the interaction effects

L_2: Sum of squares due to lack of fit of curvature

The calculation for L_2 sum of squares for estimating curvatures' lack of fit is given by Draper and Herzberg (1971), Cochran and Cox (1957, p 342), Peng (1967, p 160), and Meyer (1971, p 116) is:

$$\text{Sum of squares } L_2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)^2$$

where:

n_1 = Number of replicated center points

n_2 = Number of non-center points (fractional factorial portion)

$\bar{y}_1$ = Mean response at center points

$\bar{y}_2$ = Mean response at non-center points

L_2 has one degree of freedom and is the sum of the Bii aliased in a single string.

The L_1 sum of squares (for estimating interaction lack of fit) can be calculated as follows:

$$\text{Sum of Squares } L_1 = \begin{bmatrix}\text{Total Lack of Fit} \\ \text{sum of squares}\end{bmatrix} \text{ minus } \begin{bmatrix}L_2 \text{ sum of squares}\end{bmatrix}$$

L_1 has one less degree of freedom than the total Lack of Fit sum of squares had.

Variances are formed for L_1 and L_2 by dividing the sum of squares by the degrees of freedom. These can be tested for significance using the error term in the conventional way. If there are so few degrees of freedom in the error term of the unreplicated basic central-composite design as to make the power of such a test questionable, it would be wiser for the investigator to inspect the relative magnitudes of the proportions of variance accounted for by each of the sources of variance. (See Simon, 1976a).

Meyer (1971, p 116) shows how this technique would be used with a fractional factorial Resolution IV design augmented with center points. In his analysis (p 117), he isolated all linear model terms plus lack of fit and then error. The four degrees of the lack-of-fit term were further isolated into 3 degrees of freedom for the cross-product sources (L_1) and one degree of freedom for the quadratic sources (L_2). In this 2^{4-1}_{IV} design, the 3 degrees of freedom for the cross-product source were actually for three strings each with two two-factor interactions aliased with one another. The single degree of freedom for the quadratic source represents the sum of the coefficients of all quadratic terms.

While the wording in Draper and Herzberg's paper (1971, p 234, para 3.1) seems to suggest that this splitting of the lack of fit in a first-order model is appropriate only when the 2^{k-p} fractional factorial design is of "resolution greater than four," this is not the case. This procedure then could be used with Resolution IV screening designs to determine whether an observed lack of fit is the result of inadequate curvature of cross-product information, or both, in the first order model.

Meyer (1971, p 123) later makes an important point when he warns his readers that the aggregate sources of variance that make up the lack of fit will differ depending on the experimental design. He writes: "Essentially, they represent terms that the experimenter could have included in the model but didn't." Thus, if a lack of fit test is not significant, implying an adequate representation, the investigator should be sure that the terms of interest are included in the design. Otherwise, prediction will suffer.

SPLITTING THE LACK OF FIT OF SECOND ORDER DESIGNS

Draper and Herzberg (1971, p 235) specify that this procedure for splitting the Lack of Fits sum of squares for a second order central-composite design should be used only when the cube part of the design is Resolution VII or higher. A Resolution VII design enables all main and two- and three-factor interaction effects to be isolated from one another. In this case, L'_2 is used to check for fourth order biases.

186

Since for most psychological research valid fourth order effects are extremely unlikely (Simon, 1976b) any significant Lack of Fit of the L'_2 term would suggest that unwanted sources of variance are distorting the data.

Calculations. L'_1 will provide a test of third order biases. The calculation of L'_2 for the second order model is more complicated than for the first order model. Draper and Herzberg (1971, p 235) provide the following equation:

$$L'_2 \; SS = d(1 + dt)^{-1} \left\{ t(n-d)\bar{y}_1 - \bar{y}_o + s_i \sum_{i=1}^{k} \alpha i \right\}^2$$

The meaning of each symbol is given in Table 16. The L'_2 SS has one degree of freedom.

L_1 is obtained by subtracting the sum of squares for L'_2 from the total Lack of Fit sum of squares, thus:

$$L'_1 \; SS = (\text{Total LoF SS}) - (L'_2 \; SS)$$

The L'_1 SS has one degree of freedom less than the Total SS.

If the second-order design is orthogonally blocked, the sum of squares for blocking can be removed as usual. Since L'_1 is found by subtraction, removing the sum of squares for blocks will reduce the size of L'_1 but will not affect L'_2.

TABLE 16

SYMBOLS USED IN EQUATIONS TO CALCULATE L'_2 SUM OF SQUARES FOR THE SECOND-ORDER CENTRAL-COMPOSITE DESIGN

d = Number of center points

n = Total number of observations

k = Number of factors (independent variables)

c = Sum of non-center point coefficients squared

g = Sum of non-center point coefficients raised to 4th power

h = Sum of cross products between any pair of coefficients squared over all non-center points*

$\bar{y}_1$ = Mean performance at non-center points

$\bar{y}_0$ = Mean performance at center points

$$t = \frac{g + h\,(k - 1)}{(n - d)\,[g + h(k - 1)] - kc^2}$$

$$s = \frac{-ct}{g + h(k - 1)}$$

α_i = Sum of cross products between performance and coefficients squared of factor i over all non-center points (where $i = 1,2,\ldots k$)

---

*In the conventional central-composite design, this sum will equal the number of non-center points.

EFFECTS OF REPLICATING NON-CENTER POINTS OF THE CCD

Draper and Herzberg (1971, p 233) comment on this stating that ". . . if the center points are not the only replicated points in the design there are slight changes in the above which do not materially affect the situation." They cite some notational changes that might be made but indicate that it would not be necessary to make any changes in the calculation of L_2 or L'_2. Although L_1 and L'_1 would be affected by the change, the computations remain the same since their sums of squares is obtained by subtraction.

ADDITIONAL CRITERION FOR EVALUATING THE EQUATION

Suich and Derringer (1977, p 213) note that ". . . the significance of the regression F-ratio and the nonsignificance of the lack-of-fit F-ratio do not necessarily imply that $\hat{Y}(X)$ is an adequate [predictive] model." At best, when the regression F-ratio exceeds the critical F value for significance, this only indicates that the fitted equation is probably a better predictor of performance than the mean of the data would be. Such information is of little practical value. Draper and Smith (1966, p 64) suggest that ". . . unless the range of values predicted by the fitted equation is considerably greater than the size of the random error, prediction will often be of no value even though a 'significant' F-value has been obtained, since the equation will be 'fitted to the errors' only." J. M. Wetz (1964), a student of G. E. P. Box, in a Ph.D. dissertation, suggested that the F-ratio of the equation would have to exceed some criterion F-value by about a factor of four to be rated as a satisfactory prediction tool.

Suich and Derringer (1977) provide ". . . a numerical criterion, γ , which quantifies the range of values predicted by [a second degree polynomial] relative to the size of the standard error. That is, the importance of the standard error is considered in light of the magnitude of the changes to be estimated by the model itself . . ." (p 213). This equation is:

$$\gamma = \left[\sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{m \, \Sigma (Y_n - Y)^2} \right]^{1/2}$$

where

Y_i = Each performance score

$\bar{Y}$ = Mean performance

m = Number of terms in equation excluding the constant

n = Number of observations

Calculation and Test

Instead of wishing to compare the F-value obtained by the usual method:

$$F = \frac{\text{Regression mean square}}{\text{Error mean square}}$$

with the standard F-value taken from a central-F distribution (published in most statistics books that deal with the analysis of variance), that is, to test the hypothesis that γ is or is not greater than some non-zero value considered to be an important difference for a particular situation. To do this, they develop an equation to calculate a non-central F-value to compare with the F obtained from the experimental data. This non-central F (i.e., $F'_{\alpha,m,n-m-1,\gamma^2}$) can be estimated for any risk level, α, and particular pairs of degrees of freedom, m and (n-m-1) by adjusting the standard

F-value found in the conventional tables. This relationship is:

$$F'_{\alpha,m,n-m-1,\gamma^2} \approx (1+\gamma^2)\, F_{\alpha,b,n-m-1}$$

where:

$$b = \frac{m(1+\gamma^2)^2}{(1+2\gamma^2)}$$

and (n-m-1) is the degrees of freedom, and α is the acceptable risk level of committing a Tvpe I error (i.e., stating that a difference exists when in fact it doesn't).

The γ substituted in this equation is not calculated from the data, but is the degree of variation required for importance.* To select an F' to be approximately four times

---

*We could decide to use a γ value calculated from the data using the aforementioned equation and substitute that into the equation relating F' to F, but reversed thus:

$$F = \frac{F'}{(1+\gamma^2)}$$

with the appropriate degrees of freedom indicated above for both F and F'. Then by using the standard F-distribution tables, along with some interpolation, we could find the risk level, α, for accepting the equation as a predictor by searching the table for the F value for the indicated degrees of freedom closest to the one calculated above. One would need a set of F-tables that gives F-values for a range of probability values (e.g., Fisher and Yates, 1963).

the size of F (as Wetz had suggested), then making γ equal to 2 would roughly produce that result. However, the decision of how large this value should be is up to the investigator and a matter of experience. The experiences of the statisticians who have suggested the value might be four were not working with human performance data -- more likely it was chemical engineering data -- so we will have to try it and see how it works. Certainly any more critical criterion than the one currently in use is likely to produce a better predictive equation, although Suich and Derringer say it ". . . is not meant to be a final answer to the problem but more as a benchmark or rule-of-thumb to help in answering this difficult question. . ." (p 216).

If the regression F-value is less than F', the investigator would reexamine two things: 1) is his error variance too large because of too small a sample? 2) is the equation model adequate or should it be expanded? Both require more data to be collected. If the regression F equals or is larger than F', then we have increased our confidence in the equation as a predictive model. Suich and Derringer provide an example of this test (pp 214-216).

X. ANALYZING THE DATA FROM AN INCOMPLETE SCREENING EXPERIMENT

An experimenter may be required to do an analysis "on-line" each time a new piece of data has been collected. For example, he may wish to check his results as soon as the data is collected in order to decide whether to stop or to modify the experimental program. Or, he may wish to keep abreast of the data in the event the experiment is inadvertently terminated prematurely. While a regression analysis can be performed relatively quickly with a modern computer, it may not be convenient or may be too costly to make one available for this purpose.

Hunter (1964) has provided a "predictor-corrector" (P-C) equation that can be used to determine the regression coefficients in a polynomial model after the data has been collected on each experimental condition of a screening design (or for that matter, any 2^k and 2^{k-p} design), provided that an initial set of orthogonal estimates of the coefficients is available. This means that if a screening design is made up of blocks of Resolution III designs, then once the first block has been completed -- enabling the coefficients of a first order polynomial to be estimated -- a new equation can be determined relatively quickly after data has been collected at a new data point. The predictor-corrector equation provides an exact least squares estimate, an update, of all the coefficients without elaborate calculations or the need for a high-speed computer.

REQUIREMENTS FOR USING THE P-C EQUATION

Two conditions must be satisfied before the equation can be used:

1) The estimated coefficients from at least a single Resolution III block must be available. More, or higher resolution blocks are acceptable.

2) The rows of the new data points must be orthogonal. That means that the sum of the cross products between adjacent coefficients (i.e., plus and minus ones) of the sign matrix making up any two rows must equal zero.

Both conditions are met in a 2^{k-p} screening design made up of two Resolution III blocks. They would also be met if one Resolution IV design, to represent the initial block, had been completed and was in the process of being replicated, or a new plan begun.

PREDICTOR-CORRECTION EQUATION

The P-C equation provided by Hunter (1964, p 43) is:

$$d_i = \frac{1}{mN + q} (Y_i - \hat{Y}_i)\, r_i^{*}$$

where:

q = number of coefficients in the model; $q \leq N$

m = number of blocks of N conditions already completed

N = number of conditions per complete block

r_i = row vector of coefficients (i.e., ± 1) of

---

[*] Italicized letters are matrix symbols.

independent variables associated with the ith experimental condition

Y_i = new performance score associated with r_i

$\hat{Y}_i$ = predicted performance score associated with ith observation $(r_i B)$

The correction constants, d_i, for the ith condition is combined with the coefficients (B) from the previous block to get the revised coefficients (B^*), thus:

$$B^* = B + \sum_{i}^{n} d_i$$

The variance of each coefficient is calculated:

$$\text{Variance } (b^*) = \frac{1}{mN} \left[1 - \frac{n}{mN + q} \right] \delta^2$$

EXAMPLE

How the equation is used can best be explained by means of an illustration. Fictitious data for a 2^{2+1} fractional factorial experiment with 8 observations is given in Table XVII. Eight observations enables two Resolution III blocks of data to be collected. We will presume that the first block was run and the coefficients for the linear terms were calculated. We will use the predictor-corrector equation to obtain the least squares equation after the results from the 5th and 6th data points are each obtained. The procedure for calculating the new coefficients after each new experimental condition has been completed is as follows:

1. Calculate the q coefficients from the N experimental conditions in Block I. Yates' algorithm can be used to obtain the effects-total, which are divided by N to obtain the coefficients. The first four (N)

TABLE 17

IMAGINARY DATA WITH WHICH TO ILLUSTRATE AN INCOMPLETE ANALYSIS

	#	Exptl. Condition	(I)	A	B	C	Performance
	1	c	-	-	-	+	1.3
Block I	2	a	-	+	-	-	3.6
(I=ABC)	3	b	-	-	+	-	2.4
	4	abc	-	+	+	+	1.7
	5	ab	+	+	+	-	2.5
Block II	6	bc	+	-	+	+	1.5
(I=ABC)	7	ac	+	+	-	+	2.8
	8	(1)	+	-	-	-	3.4
	9	c	-	-	-	+	1.2

TABLE 18

WORKING DATA TO OBTAIN UPDATED EQUATIONS

$\boxed{A}$						$\boxed{C}$	$\boxed{B}$	$\boxed{D}$
		(I)	A	B	C	Y	$\hat{Y}$	d_i
		2.50	.40	-.20	-.75			
Exptl. cond.	#5	+	+	-	+	2.5	2.35	.0188
Exptl. cond.	#6	+	-	+	+	1.5	1.15	.0438
Coef. I+5		2.519	.419	-.219	-.731			
Coef. I+6		2.544	.356	-.156	-.706			

experimental conditions in Table 17 make up Block I and the four (q) coefficients for this data are shown in Table 18 at $\boxed{\text{A}}$.

2. Solve for the denominator of d_i, the correction constants:

$$d_i = \frac{1}{mN + q} (Y_i - \hat{Y}_i) r_i$$

In this example,

m = 1 block already completed
N = 4 conditions in the complete block
q = 4 coefficients in the model (including mean)

Therefore, the P-C equation for this problem reduces to:

$$d_i = \frac{1}{(1 \times 4) + 4} (\hat{Y}_i) \; r_i$$

$$d_i = \frac{(Y_i - \hat{Y}_i) \; r_i}{8}$$

3. Determine the estimated performance for the new data point, $\hat{Y}$. This is the sum of the cross products between the Block I coefficients and corresponding ±1 coefficients of the new data point. Include the plus and minus signs in this operation.

For example, in Table 17 to obtain the estimated performance for experimental condition #5, the following steps are performed:

$$\text{Coefficients}\left\{\begin{array}{lrrrr} \text{Block I:} & +2.50 & +.40 & -.20 & -.75 \\ \text{Exptl.} & & & & \\ \text{Cond. \#5:} & +1 & +1 & -1 & -1 \end{array}\right.$$

$$\hat{Y}_5 = +(+2.50)\ +(+.40)\ -(-.20)\ -(-.75) = 2.35$$

This value is located in Table 18 at $\boxed{B}$.

4. Calculate the correction constant, by subtracting the estimated performance, $\hat{Y}_i$, for experimental condition i (such as the one just calculated for experimental condition #5) from the actual performance, Y_i, (found in Table 17 and located in Table 18 at $\boxed{C}$). Divide this difference by the denominator of d_i, which was calculated in step 2:

$$\begin{array}{rl} y = & 2.5 \\ -\ \hat{Y} = & 2.35 \\ & \overline{.15} \text{ divided by } 8 = .01875 = +.019 \end{array}$$

which is the correction constant for experimental condition #5.

5. Add this correction constant (using the sign vector of the particular experimental condition) to the corresponding coefficients from the previous estimate to obtain the new estimates. These are the coefficients for the new fitted equation.

Continuing with our example:

Coefficients of previous equation (Block I)	2.50 +.40 -.20 -.75
Constant w/signs of coefficients of experimental condition #5	+.019 +.019 -.019 +.019
New equation: combined data from Block I and experimental Condition #5	2.519+.419A-.219B-.731C

6. The procedure would be repeated when performance for a new data point (#6) is obtained. The estimated performance, $\hat{Y}$, is still obtained using the coefficients from Block I. The coefficients for the new equation, however, are obtained by adding the new correction constant, multiplied by the coefficient of the corresponding columns of experimental condition #6, to the corresponding coefficients of the previous equation derived by combining Block I and experimental condition #5.

Coefficients	Block I	2.50	.40	-.20	-.75
	Exptl. Cond. #6	+1	-1	+1	+1

$\hat{Y}$ = 2.50 -.40 -.20 -.75 = 1.15

Y = 1.5

$(Y - \hat{Y})/8$ = 1.5 - 1.15 = .35/8 =+.04375 Correction Constant

Equation I+5. 2.509 .409 -.207 -.741

Constant w/ signs #6 +.044 -.044 +.044 +.044

New equation for combined data from block #1 and exptl. cond. #5 and #6 $\hat{Y} = 2.5 + .3A - .1B - .7C$

This procedure continues as each new data point is added.

If it is not necessary to estimate the equation each time a new data point is added, the correction constants along with the appropriate signs for the specific experimental conditions can be summed together and added to the original block coefficients. For example, after both experimental conditions #5 and #6 have been taken, the new coefficient for Factor A in the above example would be:

Coeff. from Original Block: .40
#5 constant: +(+.01875)
#6 constant: -(+.04375)
New coefficient: .025 (Factor A)

Computations can be made more easily when many data points are to be added if a tab with the list of correction factors (with signs) listed on it is laid next to each sign column and added or subtracted accordingly.

If a second block of N = q experimental conditions is run -- in this example, eight more -- further revisions of the equation would be based on the coefficients derived from the data from both blocks. This would also require a change in the denominator of the correction constant:

- If the number of coefficients to be estimated continued to be 4 (q), then since there are now 2 (m) blocks completed with 4 (N) conditions per block, the denominator of the correction constant would be: (2 x 4) + 4 = 12

- If the number of coefficients, q, including the mean, is expanded to 8 (which is possible with 8 independent observations), this would make the block size, N, equal to 8, now to be considered a single block. The denominator of the correction constant would be: (1 x 8) + 8 = 16.

Remember, all estimates are based on the data from the most recently completed block of a size capable of estimating all the coefficients.

MISSING DATA

At first glance it would appear that this process could be used to fill in missing data. For example, if all data points of the first block and all but one somewhere in the second block were completed, then a least squares fit of the available data made by using the P-C equation could be used to predict performance in the missing cell. In theory, this is true. In practice, for any cell of a 2^{k-p} design, the equation obtained from the Block I data would provide the _same_ estimate of a missing performance value at a point _within the experimental design_ as would an equation derived after the data from the incomplete block has been added to that of the first block. This anomoly occurs because each condition in the new block is orthogonal to the first block and therefore does not affect the original estimates.

However, the equation based on the old block data plus the data from the new incomplete block will provide _better_ estimates of data points anywhere in the experimental space _except those that are a part of the experimental design_.

XI. REFERENCES

Anscombe, F. N. and J. W. Tukey, The examination and analysis of residuals, *Technometrics*, 1963, *5*, 141-160.

Bakan, D., *On method: toward a reconstruction of psychological investigations*, Chapter One, "The test of significance in psychological research." San Francisco: Dossey-Boss, Inc., 1967, 1-29.

Beyer, W. H. (Editor), *Handbook of tables for probability and statistics*, Cleveland: Chemical Rubber Co., 1966.

Birnbaum, A., On the analysis of factorial experiments without replication, *Technometrics*, 1959, *1*, 343-357.

Bock, R. D., and E. A. Haggard, The use of multivariate analysis of variance in behavioral research. In D. K. Whitla (Ed.) *Handbook of measurement and assessment in behavioral sciences*, Reading, Mass.: Addison-Wesley, 1968, 100-142.

Booth, A. D., *Numerical methods*, London: Butterworth Scientific Publications, 1955.

Box, G. E. P., The effects of errors in the factor levels and experimental designs, *Technometrics*, 1963, *5*, 247-262.

Box, G. E. P., and K. B. Wilson, On the experimental attainment of optimum conditions, *J. Royal Statist. Soc., Series B*, 1951, *18*, 1-45.

Box, G. E. P., and J. S. Hunter, Experimental designs for the exploration and exploitation of response surfaces. In Chew, V. (Ed.) *Experimental Designs in Industry*. New York: Wiley, 1958, 138-190.

Box, G. E. P., and J. S. Hunter, The 2^{k-p} fractional factorial designs, *Technometrics*, 1961, *3*, 311-351; 449-458.

Bunde, T. A., The applications of random balance designs, *Technometrics*, 1959, *1*, 139-155.

Carney, E. J., *Ridge estimates for canonical analysis*, Ithaca, N. Y.: Cornell University, Technical Report 75-114, April 1975.

Carney, E. J., and D. A. Anderson, *A program for Monte Carlo study of ridge estimates in canonical analysis*, Ithaca, N. Y.: Cornell University, Mimeo Series, Biometric Unit, BU-520-M, June 1974.

Cattell, R. B., *Handbook of multivariate experimental psychology*, Chicago: Rand McNally, 1966.

Cochran, W. G., and G. M. Cox, *Experimental designs*, New York: Wiley, 1957 (2nd edition).

Cooley, W. W., and P. R. Lohnes, *Multivariate data analysis*, New York: Wiley, 1971.

Daniel, C., Fractional replication in industrial research. In Neyman, J. (Ed.) *Third Berkeley symposium on mathematical statistics and probability*, University of California Press, 1956, 87-98.

Daniel, C., Use of half-normal plots in interpreting factorial two-level experiments, *Technometrics* *1*(4) 1959, 311-341.

Daniel, C., Sequences of fractional replicates in the 2^{p-q} series, *J. Amer. Statist. Assoc.*, 1962, *57*, 403-429.

Daniel, C., *Applications of statistics to industrial experimentation*, N. Y.: Wiley, 1976.

Daniel, C., and F. Wilcoxon, Factorial 2^{p-q} plans robust against linear and quadratic trends, *Technometrics*, 1966, *8*, 259-278.

Daniel, C., and F. S. Wood, *Fitting equations to data*, N. Y.: Wiley - Interscience, 1971.

DeLury, D. B., *Values of the integrals of the orthogonal polynomials up to n=26*, University of Toronto Press, 1950.

Dickinson, A. W., Some run orders requiring a minimum number of factor level changes for the 2^4 and 2^5 main effect plans, *Technometrics*, 1974, *16*, 31-37.

Draper, N. R., and A. M. Herzberg, On lack of fit, *Technometrics*, 1971, *13*, 231-241.

Draper, N. R., and H. Smith, *Applied regression analysis*, N. Y.: Wiley, 1968.

Draper, R. N., and D. M. Stoneman, FActor changes and linear trends in eight-run two-level factorial designs, *Technometrics*, 1968, *10*, 301-311.

Dunnette, M. D., Fads, fashions, and folderol in psychology, *Amer. Psychol.*, 1966, *21*, 343-352.

Fisher, R. A., and F. Yates, *Statistical tables for biological, agricultural and medical research*, (Sixth Ed.) London: Oliver and Boyd, 1963.

Gallo, P. S. Jr., K. Jamieson, and K. Christian, *The strength of experimental effects in the psychological literature - How well are we doing?* Paper presented at the 57th Annual Convention, Western Psychological Association, Seattle, Washington, April 20-23, 1977.

Gnanadesikan, R., Some remarks on multivariate statistical methods for analysis of experimental data, *Industrial Quality Control*, 1963, *19*, 22-6 and 31-2.

Grossman, J. D., and H. O. Whitehurst, *Effect of visual acuity on target acquisition*, China Lake, Cal.: Naval Weapons Center, NWC-TP-5884, June 1976.

Hoerl, A. E., and R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 1970a, *12*, 55-66.

Hoerl, A. E., and R. W. Kennard, Ridge regression: applications to nonorthogonal problems, *Technometrics*, 1970b, *12*, 69-82.

Hunter, J. S., Sequential factorial estimation, *Technometrics*, 1964, *6*, 41-55.

Joiner, B. L., and C. Campbell, Designing experiments when run order is important, *Technometrics*, 1976, *18*, 249-259.

Kerlinger, F. N., and E. J. Pedhazur, *Multiple regression in behavioral research*, New York: Holt, Rinehart, and Winston, 1973.

Krane, S. A., Half-normal plots for multi-level factorial experiments. In S. S. Wilks (Chairman), *Proceedings of Eighth Conference on the Design of Experiments in Army Research, Development and Testing*, 24-26 October 1962. Washington: Walter Reed Army Institute of Research, 1963.

Lancaster, L., and S. Reynolds, *Non-randomized factorial designs characterized by trend elimination and a minimum number of factor level changes*. Falls Church, VA.: U.S. Operational Test and Evaluation Agency. Paper presented at the 21st Conference on the Design of Experiments in Army Research Development and Testing, February 1976.

Meyer, R., *Response surface methodology*, Rockleigh, N. J.: Allyn and Bacon, 1971.

Nie, N., C. H. Hull, et al, SPSS: Statistical Package for the Social Sciences (2nd Edition), New York: McGraw Hill, 1975.

Peng, K. C., The design and analysis of scientific experiments, Reading, Mass.: Addison-Wesley, 1967.

Plackett, R. L., and J. P. Burman, The design of optimum multifactorial experiments, Biometrika, 1946, 33, 305-324.

Roy, J., Step-down procedure in multivariate analysis, Ann. Math. Statist., 1958, 29, 1177-87.

Roy, S., R. Gnanadesikan, and J. Srivastiva, Analysis and design of certain quantitative multiresponse experiments, New York: Pergamon Press, 1971.

Simon, C. W., Reducing irrelevant variance through the use of blocked experimental designs, Culver City, Ca.: Hughes Aircraft Co., Tech. Rep. No. AFOSR-70-5, November 1970a, 65 pp.

Simon, C. W., The use of central-composite designs in human factors engineering experiments, Culver City, Ca.: Hughes Aircraft Co., Tech. Rep. No. AFOSR-70-6, December 1970b, 52 pp.

Simon, C. W., Considerations for the proper design and interpretation of human factors engineering experiments, Culver City, Ca.: Hughes Aircraft Co., Tech. Rep. No. P73-325, December 1971, 152 pp.

Simon, C. W., Economical multifactor designs for human factors engineering experiment, Culver City, Ca.: Hughes Aircraft Co., Tech. Rep. No. P73-326A, June 1973, 171 pp.

Simon, C. W., Methods for handling sequence effects in human factors engineering experiments, Culver City, Ca.: Hughes Aircraft Co., Tech. Rep. No. P74-451A, December 1974, 197 pp.

Simon, C. W., Methods for improving information from "undesigned" human factors experiments, Culver City, Ca.: Hughes Aircraft Co., Tech. Rep. No. P75-287, July 1975, 82 pp.

Simon, C. W., Response surface methodology revisited: a commentary on research strategy, Westlake Village, Ca.: Canyon Research Group, Inc., Tech. Rep. No. CWS-01-76, July 1976a, 60 pp.

Simon, C. W., *Analysis of human factors engineering experiments: characteristics, results and applications*, Westlake Village, Ca.: Canyon Research Group, Inc., Tech. Rep. No. CWS-02-76, August 1976b, 104 pp.

Singe, S., Personal communication, 1977.

Suich, R., and G. C. Derringer, Is the regression equation adequate? -- one criterion, *Technometrics*, 1977, *19*, 213-216.

Tatsuoka, M. M., *Multivariate analysis: techniques for educational and psychological research*, New York: Wiley, 1971.

Throndike, R. M., Canonical analysis and predictor selection, *J. Multivariate Behav. Res.*, 1977, *12*, 75-87.

Tukey, J. W., Where do we go from here? *J. Amer. Statis. Assoc.*, 1960, *55*, 80-93.

Umland, A. W., and W. N. Smith, The use of LaGrange multipliers with response services, *Technometrics*, 1959, *1*, 289-292.

Url, N., and T. Eisenberg, Predicting shrinkage in the multiple correlation coefficient, *Educat. & Psychol. Measurement*, 1970, *30*, 487-489.

Webb, S. R., Personal communication, 1977.

Wetz, J. M., *Criteria for judging adequacy of estimation by an approximating response function*, Ph.D. dissertation, University of Wisconsin, 1964.

Wilburn, Nicholas T., Application of 2^{8-4} fractional factorials in screening of variables affecting the performance of dry process zinc battery electrodes. In S. S. Wilks (Chairman), *Proceedings of Eighth Conference on the Design of Experiments in Army Research, Development, and Testing, 24-26 October 1962*, Washington: Walter Reed Army Institute of Research, 1963.

Wilk, M. B., and R. Gnanadesikan, Graphical analysis of multiresponse experimental data using ordered distances, *Proc. Nat. Acad. Sci. U.S.A.*, 1961, *47*, 1209-1212.

Wilk, M. B., and R. Gnanadesikan, Graphical methods for internal comparisons in multiresponse experiments, *Ann. Math. Statist.*, 1964a, *35*, 613-631.

Wilk, M. B., R. Gnanadesikan, and M. J. Huyett, Probability plots for the gamma distribution, Technometrics, 1962, 4, 1-20.

Yates, F., The design and analysis of factorial experiments, Harpenden, England: Imperial Bureau of Soil Science, Technical Communication No. 35, 1937.

Zahn, D. A., Modifications of and revised critical values for the half-normal plot, Technometrics, 1975, 17, 189-200.

Zahn, D. A., An empirical study of the half-normal plot, Technometrics, 1975, 17, 201-211.

Zahn, D. A., Personal communication, 1977.

APPENDICES

APPENDIX I. SUPPORT DATA FOR 2_{IV}^{8-4} SCREENING DESIGNS, N = 16

A. Three-factor interaction strings aliased to main effects . . . 192

B. Two-factor interactions aliased in strings 192

C. Inner-product sums 193

APPENDIX II. DATA FOR 2_{IV}^{16-11} SCREENING DESIGNS, N = 32

A. Screening design (sign matrix, percent overlap, factor-level change count, old factor labels, new screening labels) 194

B. Three-factor interaction strings aliased to main effects . . . 195

C. Two-factor interactions aliased in strings 196

D. Inner-product sums 197

APPENDIX III. DATA FOR 2_{IV}^{32-16} SCREENING DESIGNS, N = 64

A. Screening design (sign matrix, percent overlap, factor-level change count, old factor labels, new screening labels) 198

B. Three-factor interaction strings aliased to main effects . . . 199

C. Two-factor interactions aliased in strings 206

D. Inner-product sums 207

APPENDIX IV. COMPUTER PROGRAM FOR OBTAINING SCREENING DESIGN ALIASES . 208

APPENDIX V. PROBABILITY VALUES FOR CONSTRUCTING HALF-NORMAL GRIDS . . 215

APPENDIX VI. DERIVATION OF COMBINED LINEAR AND CUBIC TREND-ADJUSTMENT EQUATIONS 217

APPENDIX VII. HOW TO CALCULATE DETERMINANTS FOR ANALYSES FOR TWO AND THREE RESPONSE DESIGNS 218

APPENDIX VIII. ZAHN'S GUARDRAILS FOR HALF-NORMAL PLOTS 219

APPENDIX I-A

THREE-FACTOR INTERACTION STRINGS ALIASED TO MAIN EFFECTS

Original Factorial Labels	AD	ACD	AC	ABD	ABCD	ABC	AB	A
Three-Factor Interaction	ABH	ABF	ABD	ABE	BCE	ACE	ABC	ABG
Strings	ACD	ACG	ACH	ADG	BDF	ADF	ADH	ACF
Aliased with	AEF	AEH	AEG	AFH	BGH	AGH	AFG	ADE
Main Effect	BCF	BCH	BCG	BDH	CDG	CDH	BDG	BCD
	BDE	BEG	BEH	BFG	CFH	CFG	BFH	BEF
	CEH	CEF	CDE	DEF	DEH	DEG	CDF	CEG
	DFH	FGH	DGH	EGH	EFG	EFH	CGH	DFG
New Factor Main Effects	G	D	F	C	A	B	E	H

APPENDIX I-B

TWO-FACTOR INTERACTIONS ALIASED IN STRINGS

Original Factorial Labels	D	CD	C	BD	BCD	BC	B
Two-Factor Interaction	AB	AE	AC	AF	AH	AG	AD
Strings	CE	BC	BE	BD	BG	BH	BF
Aliased with	DF	DH	DG	CH	CF	CD	CG
Main Effect	GH	FG	FH	EG	DE	EF	EH

---

[*] For experimental design, Table 1, page 15 in text.

APPENDIX I-C

2^{8-4} Design — INNER-PRODUCT SUMS* (TX) — N=16

			Screening Factor Labels								Two-Factor Interaction Labels						
			A	B	C	D	E	F	G	H	AH	AG	AF	AE	AD	AC	AB
TRENDS	Linear	LX								16					32	64	128
	Quadratic	QX					16	32	64			64	128	256			
	Cubic	CX		320	640	1280				672	2560				1264	1888	-1344
Original Factorial Labels			ABCD	ABC	ABD	ACD	AB	AC	AD	A	BCD	BC	BD	CD	B	C	D

*Percentage overlap between trend and factor = $r^2 \times 100$; $\sqrt{r^2}$ = correlation

$r^2 = \frac{(TX)^2}{N(TT)}$ where TX = LX or QX or CX, the inter-product sum between trend and factor and TT = LL or QQ or CC, the trend coefficient squared.

For N = 16, LL = 1360, N(LL) = 21760; QQ = 5712, N(QQ) = 91392, CC = 1007760, N(CC) = 16124160

Where cell has been left blank, the inner-product sum is zero.

PAGE 194 REMOVED FROM DOCUMENT PRIOR TO
BEING SHIPPED TO EDRS FOR FILMING.
OVERSIZED PAGE.

APPENDIX II-B

THREE-FACTOR INTERACTION STRINGS ALIASED TO MAIN EFFECTS

NEW MAIN EFFECTS	ORIGINAL FACTOR LABELS	ALIASED THREE-FACTOR INTERACTION STRINGS														
O	AE	ABP CFP HLP	ACK CGI IJN	ADJ CHJ IKM	AEH CLM JNP	AFN DFI KNP	AGM DGP	AIL DHK	BCN DLN	BDM EFG	BEL EIP	BFK EJK	BGJ EMN	BHI FHM	CDE FJL	GKL GHN
K	ADE	ABN CFN HLN	ACO CGM IJP	ADE CIL IMO	AFP DFM JMN	AGI DGN NOP	AHJ DHO	ALM DLP	BCP EFL	BDI EIN	BEG EJO	BFO ENP	BHM FGJ	BJL FHI	CDJ GHP	CEH GLO
N	AD	ABK CFK HKL	ACP CGJ IJO	ADI CHI IMP	AEG DFJ JKM	AFO DGK KOP	AHM DHP	AJL DLO	BCO EFH	BDE EIK	BFP EJP	BGI EMO	BHJ FGM	BLM FIL	CDM GHO	CEL GLP
J	ACE	ABM CGN HLM	ACE CHO IKP	ADO CLP INO	AFI DEH KMN	AGP DFN MOP	AHK DGM	ALN DIL	BCI EGL	BDP EIM	BEF EKO	BGO ENP	BHN FGK	BKL FHP	CDK FLO	CFM GHI
F	ABC	ABC CHL HJP	ADL CJM HMO	AEM CKN ILN	AGH COP JLO	AIJ DEP LMP	AKP DIO	ANO DJN	BDH DKM	BEJ EGO	BGL EHN	BIM EKL	BKO GIP	BNP GJK	CDG GMN	CEI HIK
L	AB	ABH CIK GKO	ACG CJP GNP	ADF CMO HJM	AEP DEM HKN	AIO DGH HOP	AJN DIJ	AKM DKP	BCD DNO	BEO EFK	BFG EGJ	BIP EHI	BJK FIN	BMN FJO	CEN FMP	CFH GIK
P	A	ABO CFO HLO	ACN CHM IJK	ADM CJL IMN	AEL DEF JMO	AFK DGO KNO	AGJ DHN	AHI DKL	BCK EIO	BDJ EJN	BEH EKM	BFN FGI	BGM FHJ	BIL FLM	CDI GHK	CEG GLN
E	ACDE	ABI CGP IOP	ACJ CHK JKO	ADK CLN JNP	AFM DFP KMP	AGN DGI MNO	AHO DHJ	ALP DLM	BCM FGO	BDN FHN	BFJ FKL	BGK GHM	BHP GJL	BLO HIL	CDO IJM	CFI IKN
I	ACD	ABE CGO GLM	ACM CHN JKP	ADN CKL JNO	AFJ DEG KMO	AGK DFO MNP	AHP DHM	ALO DJL	BCJ EHL	BDK EJM	BFM EKN	BGN EOP	BHO FGP	BLP FHK	CDP FLN	CEF GHJ
M	AC	ABJ CGK HJL	ACI CHP IKO	ADP CLO INP	AEF DEL JKN	AGO DFK JOP	AHN DGJ	AKL DHI	BCE EGH	BDO EIJ	BFI EKP	BGP ENO	BHK FGN	BLN FHO	CDN FLP	CFJ GIL
H	ABE	ABL CIN GKP	ACD CJO GNO	AEO CMP JLM	AFG DEJ KLN	AIP DGL LOP	AJK DIM	AMN DKO	BCG DNP	BDF EFN	BEP EGM	BIO EIL	BJN FIK	BKM FJP	CEK FMO	CFL GIJ
D	ABDE	ABG CIP HKP	ACH CJK HNO	AEK CMN ILM	AFL EFP KLO	AIN EGI LNP	AJO EHJ	AMP ELM	BCL FIO	BEN FJN	BFH FKM	BIK GHL	BJP GJM	BMO GKN	CEO GOP	CFG HIM
C	ABCE	ABF DIP HJO	ADH DJK HMP	AEJ DMN IKL	AGL EFI JLP	AIM EGP LMO	AKO EHK	ANP ELN	BDL FHL	BEM FJM	BGH FKN	BIJ FOP	BKP GIO	BNO GJN	DEO GKM	DFG HIN
A	ABCDE	BCF DIN HJK	BDG DJO HMN	BEI DMP ILO	BHL EFM JLN	BJM EGN KLM	BKN EHO	BOP ELP	CDH FGH	CEJ FIJ	CGL FKP	CIM FNO	CKO GIK	CNP GJP	DEK GMO	DFL HIP
B	ABCD	ACF DIK HJN	ADG DJP HKM	AEI DMO ILP	AHL EFJ JKL	AJM EGK LMN	AKN EHP	AOP ELO	CDL FGL	CEM FIM	CGH FKO	CIJ FNP	CKP GIN	CNO GJO	DEN GMP	DFH HIO
G	ABD	ABD CIO HKP	ACL CJN HNO	AFN CKM ILM	AFH DEI KLO	AIK DHL LNP	AIP DJM	AMO DKN	BCH DOP	BEK EFO	BFI FHM	BIN EJL	BJO FIP	BMP FJK	CDP FMN	CEP HIJ

APPENDIX II-C

TWO-FACTOR INTERACTION ALIASED IN STRINGS

	ORIGINAL FACTORIAL LABELS														
	E	DE	D	CE	CDE	CD	C	BE	BDE	BD	BCE	BCDE	BCD	BC	B
STRINGS OF ALIASED TWO-FACTOR INTERACTIONS (NEW LABELS)	AB CF DG EI HL JM KN OP	AF BC DL EM GH IJ KP NO	AC BF DH EJ GL IM KO NP	AG BD CL EN FH IK JP MO	AL BH CG DF EP IO JN KM	AH BL CD EO FG IP JK MN	AD BG CH EK FL IN JO MP	AI BE CM DN FJ GK HP LO	AM BJ CI DP EF GO HN KL	AJ BM CE DO FI GP HK LN	AN BK CP DI EG FO HM JL	AP BO CN DM EL FK GJ HI	AO BP CK DJ EH FN GM IL	AK BN CO DE FP GI HJ LM	AE BI CJ DK FM GN HO LP

2^{16-11}_{IV} Design APPENDIX II-D. INNER-PRODUCT SUMS* N = 32

Screening Factor Labels

TRENDS			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Linear	LX																32
	Quadratic	QX												32	64	128	256	
	Cubic	KX						128	256	512	512	1024	2048					1088
Original Factorial Labels			ABCDE	ABCD	ABCE	ABDE	ACDE	ABC	ABD	ABE	ACD	ACE	ADE	AB	AC	AD	AE	A

Two-Factor Interaction Labels

		AP	AO	AN	AM	AL	AK	AJ	AI	AH	AG	AF	AE	AD	AC	AB
	LX												64	128	256	512
	QX						128	256	512	512	1024	2048				
	KX		1024	2048	4096	8192							2144	4032	6016	-4352
		BCDE	BCD	BCE	BDE	CDE	BC	BD	BE	CD	CE	DE	B	C	D	E

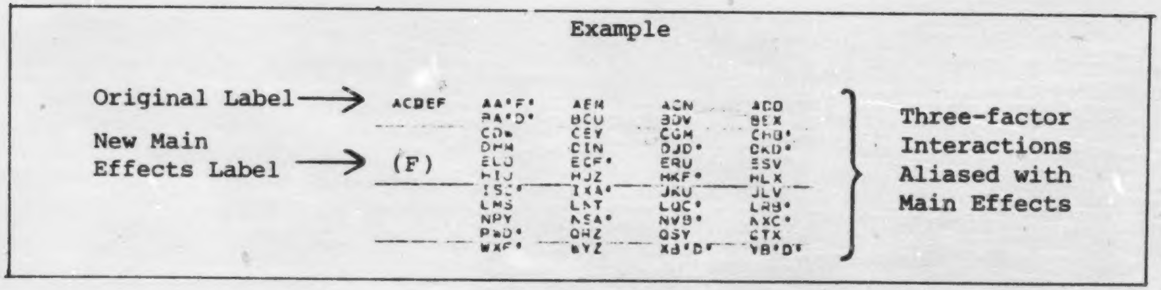
---

*See footnotes in Appendix I-C.

PAGE 198 REMOVED PRIOR TO BEING SHIPPED
TO EDRS FOR FILMING. OVERSIZED PAGE.

APPENDIX III-B

THREE-FACTOR INTERACTION STRINGS ALIASED TO MAIN EFFECTS


Example
Original Label
New Main Effects Label
ACDEF
(F)
AA'F' BA'D' CD DHM ELO HIJ ISC' LMS NPY PWD' WXF'
AEM BCU CEY CIN ECF' HJZ IKA' LNT NSA' QHZ WYZ
ACN BDV CGM DJD' ERU HKF' JKU LOC' NVB' QSY XB'D'
ADO BEX CHB' CKD' ESV HLX JLV LRB' NXC' CTX YB'D'
Three-factor Interactions Aliased with Main Effects

ACDEF	NC'D'	LUA'	LCC'	LRE'	LNT	LMS	YB'D'	ZB'C'	KOT	KMR	NSA'	KSB'	KQD'	UD'D'	KVA'	JQD'
(F)	JTE'	JLV	JKU	UZF'	JWA'	KLW	JOS	JNR	PQA'	BHO	UVW	BGN	BDV	BEX	OPZ	QTX
	NEY	CJC'	CKP	XB'D'	MPX	CLD'	UXY	MTA'	CHB'	CIO	CEY	CGM	ORA'	BA'D'	CDW	BIE'
	BJP	STU	ESZ	BTF'	BSY	BQW	BKC'	BLD'	ISC'	IPT	IRD'	PUC'	ILY	OC'F'	[illegible]	IXA'
	IJF'	IKZ	HYA'	HTC'	PVD'	HRD'	HPS	HLX	HNQ	NOW	HJZ	EKF'	HIU	OYD'	[illegible]	PWD'
	GIV	GHW	GSD'	GTD'	WXF'	GLF'	WYZ	GJY	GKX	NVB'	GZA'	MD'F'	GOQ	GPR	VC'D'	EA'B'
	NXC'	ESV	ETW	ERU	EQF'	OXD'	MZD'	PB'F'	EJM	EID'	EKN	MYC'	ELO	EGC'	EHD'	NZD'
	VYF'	DA'C'	BTV	DIN	DHM	DKD'	DLP	DJD'	DQU	DTY	QRZ	DRF'	RSN	DSX	MOV	VXZ
	CA'D'	DGB'	DEZ	MWB'	QSY	CQV	OUB'	CSF'	CTZ	CRX	ND'F'	AGO	AEP	ADN	AQB'	AIW
	AJX	ATE'	ECU	AA'F'	ACN	ADO	ABC'	ASD'	AHV	ALZ	AKY					

ABCDEF	MFY	OQU	NCV	PRU	PQF'	BDH	BEJ	BFM	BOV	BPX	WRX	BWD'	BTA'	MSZ	MQW	BKE
(A)	BIJ	BNU	ELS	BZD'	BYC'	CEK	BD'F'	CDI	CFN	CHQ	CJR	ORF'	CNG	CLT	MTF'	LQR
	PSV	LUF'	IVX	OSX	LWY	JMP	IC'D'	JNC'	JOD'	QXD'	JQT	WSF'	JVZ	JUY	JWF'	QYE'
	OTY	JB'D'	FTW	KNP	QZC'	KMC'	KOD'	KQS	KVF'	KWZ	KB'D'	KUX	LHE'	LOP	LND'	CPY
	COW	CSA'	CVE'	DFO	DGQ	CZD'	CD'F'	DEL	CXC'	DRA'	DUB'	DJS	DKT	DMV	DNW	MA'D'
	DPZ	EFP	IC'F'	DYD'	DXD'	EHS	EGR	EOZ	ENY	EMX	EIT	EQA'	EB'F'	FGU	FHV	PA'B'
	FJX	FKY	FIN	EUC'	EVD'	EWD'	MA'D'	GVW	GD'D'	GZF'	GXY	GST	GOB'	GPC'	GMN	GLA'
	GHI	FA'F'	GJK	RVD'	FTD'	FLZ	FQB'	FSD'	FRC'	MKA'	LB'C'	BUD'	HJL	HMO	OA'C'	HNE'
	HPD'	HET	HYF'	MDW	HXZ	IKL	HC'D'	IJA'	IMB'	INO	IPD'	IRS	IYZ	IXF'	RZB'	NTZ
	IUV	TUD'	SYB'	VYA'	SUD'	WXA'	SWC'	UZA'	TXB'	TVC'	BCG					

ADEF	OQB	BGY	BDD'	BEM	BFJ	NQS	XZD'	XYC'	CDD'	CEN	UWD'	CFK	CGX	CJU	CHF'	CIZ
	CLW	CMB	YZE'	BQD'	BOS	BWA'	BTB'	EKU	BHZ	BIP'	UVD'	BQT	BLV	XD'F'	LTY	LRF'
(P)	LSX	LQU	JID'	JKC'	JFY	JSZ	JQW	JTP'	ZC'F'	JA'D'	MUY	TXA'	UZB'	KA'D'	YD'F'	KLL'
	KSF'	KTZ	MVZ	KQV	KRX	VWC'	COT	CVA'	CSB'	CQD'	DPL	DGF'	OUF'	DEO	OVX	DEB'
	DUA'	DQC'	OWY	DNT	DHS	DKW	DIY	NB'E'	MWF'	DJV	DHX	HOD'	RSD'	VYB'	EHV	EGU
	EKY	EJX	EIW	ELZ	HTD'	FHS	FGR	EQB'	MNC'	ERC'	ESD'	FIT	SYA'	ETD'	EA'F'	GOA'
	NVP'	GLB'	GJN	GKM	NWZ	GQZ	WXB'	GSW	GTV	FE'F'	GID'	GHD'	FQA'	FOZ	FMX	MB'D'
	PNY	FVD'	PUC'	FWD'	OB'C'	HTU	HIC'	HLM	HKB'	HJO	HNA'	LA'C'	HRW	EQY	STC'	EZA'
	C'EE	IKO	IJE'	ILN	NCD'	NUX	IHA'	IQX	ISU	IRV	ADZ	AGC'	AJH	BCC'	AA'B'	ASV
	ABU	ATW	AIE'	AMD'	ALO	AKW	AQF'	ABX	ACY	AEF	BNR					

ABCDE	MFC'	MQB'	CLA'	CMN	CJK	CHI	ORD'	CZN	CFU	CEQ	MSD'	PQD'	NTD'	LUD'	LPV	LWC'
	LCX	LNP'	LRZ	KZE'	KWD'	KVD'	KPU	OTC'	KNX	KOP'	KMY	JB'F'	WSD'	JVD'	JWD'	KLQ
(B)	PTB'	JUC'	JCA'	JOZ	JNY	LYB'	JMX	IXD'	IYD'	NTD'	IZC'	IPF'	NPR	INV	IOU	INW
	IJT	ILR	HC'F'	IKS	HXD'	HYD'	HUB'	HBA'	RPZ	MNW	MMV	HJS	RKT	PWA'	FSZ	FAY
	FGB	FLD'	RA'F'	FTF'	GRQ	FA'E'	GJR	GLT	GPY	GMU	GOW	GSA'	GXC'	GZD'	NOQ	GD'F'
	GVB'	EWF'	OYA'	EVZ	EUY	FIB'	FKC'	FJP	FHO	FGN	EB'D'	EQT	EIA'	ENC'	EOD'	EMP
	EGK	NZA'	EHL	DXZ	DYF'	DC'D'	MFX	QRS	DPD'	DNB'	DKA'	DMO	DJL	DUW	DRT	CXY
	DES	CD'D'	CZF'	DGI	DEV	CVW	CST	OPS	CPC'	COB'	TWX	UA'D'	TVY	QXF'	QYZ	XA'B'
	CUV	RB'E'	SVX	SUF'	RWZ	RVP'	SWY	RUX	VA'C'	QC'D'	TUZ	SB'C'	AWB'	AD'F'	AYC'	ATA'
	AIQ	ALS	AKU	ACG	AEJ	ADH	AOV	APX	AFM	AZD'	AKR					

ABCDF	BYU	BHI	BER	QUW	BEQ	NQB'	BD'D'	ORD'	RB'D'	BZF'	BMN	BLA'	BJK	MSD'	BXY	BST
	EVW	EPC'	ECE'	PQD'	IXD'	NTD'	IYD'	IUB'	IRA'	IPZ	INW	IMV	IJS	PVA'	IKT	HXD'
(C)	SUZ	HZC'	HYD'	HPF'	HOU	HNV	HMW	HLR	RWF'	SVY	HJT	HKS	FRX	PQV	FLD'	FTZ
	FSP'	SWX	FA'D'	GIQ	GKR	GLS	RVZ	GPX	GOV	GNU	GTA'	GWB'	GZD'	GYC'	GD'F'	EWZ
	EVP'	EUX	FUY	FXO	PJC'	FKP	FHB'	RB'E'	FGN	EQS	EOD'	OXA'	NZA'	EMC'	ENP	EGJ
	KCQ	ZIL	EHA'	DYZ	EXP'	DC'D'	QBT	EFY	DNO	DPD'	DHB'	DKL	WA'F'	DJA'	DUV	DRS
	DET	EGH	EFW	QC'D'	MFR	OPT	LWZ	LMF'	TUF'	LOY	TVX	KUC'	KB'F'	PSB'	KVD'	KWD'
	KQA'	KCZ	KPX	LXB'	TWY	KNY	QYF'	JVD'	JWD'	NSD'	JZB'	JOF'	JPU	JNX	JMY	OSC'
	IC'F'	JLQ	CXZ	WBC'	LVC'	LUD'	LPW	NTD'	UA'D'	WA'C'	YA'B'	TB'C'	ABG	AZD'	AVB'	ASA'
	AXC'	AHQ	AMU	ALT	AEI	AEK	AJR	AFN	AD'F'	AOW	APY					

ABCEF	TB'D'	MRD'	FCC'	EGI	BEV	BES	CET	CPW	CGH	PRB'	CKL	CJA'	BC'D'	BYF'	ORC'	BNB'
	BKA'	HMO	QVW	BJL	BUW	BXZ	BRT	OQB'	EPD'	QXI	LUC'	LVD'	MTC'	LQA'	LMX	LNY
(D)	LWD'	LOZ	NFD'	KXE'	KUD'	KVC'	OTD'	KPW	ZA'B'	QZF'	TVF'	NSC'	KOY	KMF'	TWZ	KNZ
	JUC'	JWC'	JYB'	JPV	JNF'	JSZ	OSD'	JOX	ID'F'	JKQ	IXC'	NPT	PUA'	IZD'	IVB'	ISA'
	IPY	WA'E'	ILT	IOW	IMU	MD'F'	SUY	IJR	HYC'	HZD'	OA'F'	HWB'	MTA'	HPX	HKE	SVZ
	EIQ	HLS	HNU	HOV	FFF'	FSX	FQU	FLP	RWX	FTI	MNQ	FA'C'	GJI	GKS	GLE	GEW
	GEV	SWF'	RVY	GOU	GPF'	LB'F'	GXD'	GYD'	GZC'	EUF'	EWY	EVX	RUZ	NXA'	FKD'	VA'D'
	FJD'	FGB'	EE'C'	PIN	FUM	EQB	ZOP	EMD'	END'	EBJ	EGA'	MYA'	EIK	SB'D'	EFZ	TUX
	CIZ	CXF'	CC'D'	QD'D'	QST	CNO	CRB'	CPD'	CUV	CRS	MPS	ACI	AEL	BCQ	ARA'	APO
	AFZ	AUB'	AC'F'	AYD'	ABN	AGQ	AKT	AMV	ANN	AJS	AXD'					

ABDEF	LUB'	YA'C'	TYZ	TXF'	LPZ	LRA'	LNW	KD'F'	LMV	OUA'	KZD'	KXC'	PSD'	KVB'	KPY	ZA'C'
	KSA'	KMU	KCW	NSB'	JNU	JPX	JOV	JLS	JKR	JD'F'	KLT	LXD'	JYC'	JWB'	JZD'	JTA'
(E)	PTE'	TC'E'	RZF'	PQB'	BHL	BFX	BGK	BDS	MQD'	CIL	CHA'	QWX	WQD'	CGJ	CPY	XA'D'
	PBC'	MTB'	BE'D'	ORB'	CDT	RD'D'	DIA'	BWF'	EUY	BVX	QUZ	OQC'	BQT	BOD'	QVY	BKC'
	BMP	IPW	SUW	IJQ	ICY	INF'	INZ	IUD'	IVC'	IXB'	LYD'	NYB'	MUD'	MWC'	HPV	HNF'
	HOX	BXY	WMZ	HIR	HAQ	FQF'	FRU	FSV	FTW	MNR	GHT	GIS	FA'B'	GLQ	GOF'	GPU
	CNX	GMY	IC'F'	SXZ	NVA'	RVW	GWD'	GVD'	NOT	GZB'	FLO	FID'	FJM	FKS	PGC'	FHD'
	SYF'	PA'F'	MOS	DB'C'	DIX	DNJ	TUV	DMD'	DQR	DOP	DWD'	SC'D'	DUF'	DWY	DVX	CVF'
	CWZ	CB'E'	RST	EGA'	DPZ	CMC'	CNP	COD'	CQS	CUX	MWA'	AFP	AB'F'	BCR	ADL	ABJ
	AVD'	AIT	ACZ	ANY	AGR	AMX	AWD'	ACK	AUC'	AQA'	AHS					

ABDF	ELA'	TC'F'	OEV	BEG	BFC'	NSW	BHT	MQZ	GQX	BNX	BMY	BLQ	BOF'	BPU	BVD'	ED'F'
	BED'	NEU	BIS	QUD'	MTV	CDL	UWA'	BZB'	PEX	CMX	OSU	NQF'	CGR	CFP	CHS	CIT
(K)	LOV	TYD'	YA'F'	LPD'	LRS	PQV	XZA'	QWC'	LXF'	JST	QYB'	JVW	JZF'	PTZ	JXY	JD'D'
	JIA'	JMN	TXE'	IB'C'	NTL	JOB'	LYZ	JPC'	PSF'	LNO	NSY	LMB'	CUC'	CQA'	IUB'	COZ
	CNY	DGS	DFD'	CB'F'	CVD'	CWD'	DEI	DUD'	DVC'	NA'B'	DNZ	DOY	DPW	DMF'	MOA'	DJC
	EBR	EXB'	EFN	EHQ	ELT	EJR	EOW	RTA'	EMU	EPY	SZC'	ESA'	ED'F'	FHF'	FGX	FA'D'
	SYD'	FJU	FIZ	EXC'	FZD'	SXD'	EVB'	GVZ	GWF'	GB'D'	HIJ	GUY	GQX	LC'D'	GNC'	GOD'
	GMP	GIA'	GEL	FVA'	PSB'	RWB'	FQD'	FOT	FLW	FMR	HND'	NMD'	HOC'	HPB'	HWX	HVY
	RUZ	RZD'	INY	A'CD	IVX	IUF'	IOP	IND'	RYC'	IMD'	IQR	AUX	BCJ	AWZ	AIL	APY
	ADT	ACE	AER	AGJ	ABA'	AQS	AOD'	ANP	AMC'	AB'D'	AVF'					

ABEF	MFW	PSX	QWE'	PQU	JOW	JRT	JYF'	JXZ	JC'D'	PTY	OTW	JKA'	JMO	IB'D'	TXC'	JNB'
	JFD'	TVB'	KC'D'	KYZ	KXE'	KUV	KPD'	KRS	ZA'F'	QZB'	KMB'	KNO	NSU	ORU	TZD'	XYA'
(L)	BEJ	BEH	TE'F'	BGT	BFD'	NQX	BMZ	BNF'	PRF'	BOX	BPV	OQP'	MTU	BUD'	BWC'	NRV
	MCY	EIR	EKQ	QVD'	CDK	BYB'	CEI	CFD'	CGS	CHR	CJQ	OSV	CUD'	CHF'	VWA'	COY
	CPW	CNZ	DSF	DGR	CVC'	CXB'	DVD'	DUC'	DQA'	DNY	DOZ	SD'F'	DMX	DIT	DHS	DB'F'
	EWD'	EFO	EGQ	EKT	SZD'	EJS	ENW	EMV	ERA'	EPZ	MNA'	FGF'	PA'C'	EC'F'	FHX	FKN
	FJV	FIY	SYC'	EUB'	EXE'	EYD'	GVY	GWX	GUZ	GOC'	GMD'	GND'	GPB'	SWB'	GHK	GIJ
	FUA'	FQC'	FEB'	FMS	FNT	HIA'	OA'B'	HMP	HNC'	HOD'	HQT	HWF'	HUY	HVZ	HXD'	HB'D'
	A'EL	STA'	INZ	RZC'	IVF'	IUX	RYD'	INP	IMC'	IOE'	IQS	AB'C'	BCA'	AVX	AUF'	AIK
	AGA'	ACT	AES	AHJ	AND'	AND'	AFZ	AWY	AOP	AQR	ADE					

ACDE	ORY	QSC'	BDO	EEP	BGU	BHV	BLZ	BKY	BQB'	BRC'	UVB'	BTD'	BSD'	BJX	YZB'	BIW
	EA'F'	CEB'	CEC'	CFG	UXC'	CIV	CHW	CLP'	CKX	CJY	XB'F'	LQY	LRU	LTU	QXA'	OPD'
(M)	NPC'	UZD'	JA'B'	JLO	JKN	JTW	JRU	JCF'	JSV	UE'F'	OUW	KQZ	KTV	KSW	KLB'	KOA'
	LMA'	NST	CZD'	CSD'	CFR	COQ	DED'	DFH	ND'D'	DGW	QRD'	CZA'	DRD'	DTC'	NZF'	DNQ
	RSB'	DPS	VXE'	DLX	DKP'	DJZ	DIU	VYD'	EYA'	EFJ	EHZ	EGY	EKU	EIF'	ENR	ELV
	EB'D'	EOS	CXZ	EQD'	FKR	FIQ	NXY	ETB'	EWA'	WXD'	VC'F'	HIN	GA'E'	RVA'	GLD'	GJC'
	GKP	GTZ	GSF'	GQV	GEX	FD'F'	GIO	GHB'	WZC'	FPX	WYD'	FLS	PWF'	FOV	FNU	FWE'
	FZD'	FYC'	OYF'	FTA'	HKD'	HJD'	HLP	NVW	HRF'	HSX	HQU	NOB'	HTY	SUA'	IKD'	HA'C'
	PVZ	IJD'	ITX	PUY	PCT	ILC'	OC'D'	IRZ	B'CD	IPA'	ISY	ABF	AKC'	AQW	AEX	AJP
	AEC	ASZ	AA'D'	ACU	BCN	ABY	AIB'	AGN	ALD'	ATF'	ADV					

ACDF	QTE'	OPD'	XZE'	WC'F'	BEC'	BFG	BDB'	MST	BHW	BLF'	BKX	BPR	BOQ	BTD'	OBS	BSD'
	BJY	BIV	CEO	UWB'	BZA'	CGU	CEP	MPC'	CIW	CHV	CKY	CJX	CLZ	LSU	LRV	LQX
(N)	JCA'	KLO	UZE'	JKM	UYC'	PUX	JLB'	JTV	JQZ	JSW	QYA'	OUV	UD'F'	KA'B'	KQP'	KRU
	KTU	KSV	MVW	LMA'	YE'F'	CQB'	CSD'	CTD'	CRC'	DFI	DED'	DGV	TUA'	CA'F'	QRD'	DSC'
	DEC'	DPT	EJF'	VXD'	DKZ	DLY	DMQ	NXY	EHU	MOB'	VYD'	DXA'	EFK	EHF'	EGX	VZC'
	EIZ	RTB'	EJU	ELW	EME	EOT	PB'D'	FUQ	ESB'	ECD'	OYZ	OXF'	FJR	MZF'	EVA'	WXE'
	ME'E'	GA'E'	EIM	GKC'	GJF	GLD'	GTF'	GSZ	GRY	GQW	GHO	GIB'	FD'F'	FPY	RWA'	FSA'
	FCW	WYE'	FLT	FMU	FZD'	FXC'	FVB'	HKD'	HJD'	PWZ	HLC'	HSY	PVF'	HPA'	HRZ	HIX
	IKD'	ILP	IJD'	OC'D'	ITY	PQS	IA'C'	B'CD	IQU	ISX	IRF'	ASF'	AQV	ALD'	AHB'	AJC'
	AKP	ATZ	AA'E'	BCM	ACF	ADW	AEU	ARX	AGM	AEY	AIO					

ACEF	UXD'	LTW	LCF'	LRU	LSV	JWA'	UZC'	KLN	QZA'	JKB'	UYD'	MUW	JLM	JQY	JRW	JTU
	TVA'	KMA'	ZE'F'	KSU	KEV	KQX	NPD'	MRT	BDM	BFH	BED'	BGW	QTD'	XYB'	QSD'	WD'F'
(O)	BKF'	BLX	PRE'	BPS	BTC'	BIU	BJZ	CDN	EYA'	WRS	CGV	CED'	CFI	MPD'	CHU	CKZ
	CJF'	CLY	CEQ	CPT	VWB'	CRD'	CSC'	DGU	WC'D'	DEP	CXA'	DTD'	DSD'	DRC'	DQB'	DKY
	DIZ	MXZ	EIW	DHV	DJX	EFL	VYC'	DA'F'	EHX	EGF'	PB'C'	EIY	EJV	EKW	MYF'	ENT
	EMS	VZD'	NXF'	FGQ	ERB'	MNB'	EQC'	NYZ	FKT	FJS	EUA'	GPA'	WXC'	VD'F'	GJD'	GKE'
	GLC'	GTX	GSY	GRZ	GHN	FC'F'	SWA'	MC'D'	GIM	FRA'	FPZ	FMV	FNW	FYD'	FXD'	FUP'
	EIE'	HKC'	HJP	MLD'	PWY	LA'B'	HSZ	HRY	WZD'	HQW	HTF'	HA'D'	PVX	IJC'	IKP	ITZ
	STB'	PQR	IA'D'	B'DD	PUF'	NUV	ILD'	IRX	ISF'	IGV	BCB'	ARF'	ATY	ASX	AGB'	AEZ
	ACW	ABV	ACU	AKD'	ALP	AJD'	AHM	AIN	AA'C'	ADF	DNQ					

AEF	BWA'	UWX	BLM	BKB'	BIC'	BJO	BTU	BRW	OQA'	BQY	UVY	TA'D'	XYF'	BPW	BDX	BEV
(Z)	DHF	BGD'	PFS	CFW	CFT	MPV	CLN	CKO	CJB'	CIP	CHC'	CGD'	PEA'	IC'D'	CDY	QIV
	LCB'	LRC'	LSE'	LTD'	KPT	KRD'	KSC'	KMQ	YC'D'	KXA'	JNQ	JPS	MUD'	JYA'	KLY	JED'
	JTC'	JLX	JKF'	ISB'	ICD'	ILW	IOT	STF'	IMR	IJU	OB'F'	PUB'	IVA'	WB'D'	HQD'	NOY
	HCS	LA'F'	PLV	HWR	HIF'	HKU	HTB'	D'DF	NUD'	HWA'	FUF'	FVX	FWY	FOP	FND'	FMD'
	FCB	FB'C'	GJX	GHY	GKV	GJW	PXD'	GLU	GMT	GMS	GOR	GPQ	BVC'	PYD'	ETY	EA'C'
	VB'D'	NXB'	EFF'	EQU	ESX	FGA'	FIK	FHJ	MNF'	EKD'	ELP	EJD'	MYB'	EC'F'	EHM	OWE'
	EGE'	SA'D'	EIN	RTX	DA'B'	DID'	DHD'	DJM	RSY	DKN	OVD'	DLO	ROX	DQF'	DRU	DSV
	DTW	MWC'	DEF	VWF'	DGC'	QSW	OUC'	CMA'	CQX	CSU	CRV	ARB'	AIY	ANX	ANT	AMS
	AKW	AGF'	AUA'	ECF'	AFL	AEQ	ACD'	ADP	ABD'	AOC'	AJV					

ABCD	BMU	ELT	MQV	BJR	NRY	BPY	BOW	BXC'	MSF'	BVB'	BSA'	NQU	OPA'	BEK	BDI	BPN
	BHQ	CDH	OFZ	EE'F'	CEJ	BZD'	CFM	BTZ	CIQ	CKR	CLS	MBX	LWX	LVY	OSY	LOZ
(G)	LFB'	IC'E'	JOE'	JQS	JHC'	JMP	JUX	JWZ	JVF'	OTX	JB'D'	KOD'	KNC'	KMP	NSZ	KQT
	KVZ	KWF'	KUY	KB'D'	LMD'	LND'	LOC'	CPX	COV	CMU	CTA'	DFB'	DEA'	CD'F'	CYC'	CZD'
	CLB'	DJT	XA'D'	DLE	DMW	DKS	DOU	DNV	DPF'	DZC'	DYD'	DXD'	EIS	EHT	EPC'	ENX
	EFY	ELQ	EPU	EOF'	FBW	FJY	FKX	FIV	EVD'	EWD'	MA'D'	EZX'	FTD'	FZA'	FLF'	FOQ
	FSD'	FPR	HKL	RJA'	HEB'	HNO	HPD'	HRS	HYZ	HXF'	HUV	HC'D'	IKA'	IJL	IMO	INE'
	KTF'	IPD'	IFT	IYF'	IUW	IXZ	VXA'	QC'F'	PUC'	QRA'	RVD'	WYA'	SVC'	SXB'	RWD'	SUD'
	BE'F'	PQZ	QUE'	TWC'	QYD'	QXD'	A'BC	TUD'	UA'F'	PTV	TXB'	PSW	ABC	ADQ	AFU	ALA'
	APC'	AXY	AST	AZF'	AE'D'	AOB'	AER	AVW	AMN	AHI	AJK					

ABCE	PQY	QSA'	TZB'	MRF'	BEL	PRW	BFO	ORY	MSX	BGQ	BPZ	BRA'	DXD'	BUB'	BKT	BJS
(H)	BCJ	BNW	OCW	BMV	NBZ	QVB'	BYD'	CDG	EC'F'	CKS	CLR	CMW	CJT	ETY	UXA'	CEB'
	CZA'	NPA'	CSZ	LVZ	LUY	LQT	ID'D'	QXC'	JMD'	NTX	JOP	JND'	JQR	TWD'	JWY	JB'C'
	JVX	JUF'	PTU	KAD'	QZD'	KMD'	KOC'	TVD'	KPB'	NSY	KWX	OTF'	KVY	KUZ	LWP'	LOD'
	LNC'	LMP	CPF'	COU	CWV	DFS	QD'F'	EEJ	CXD'	CZC'	CYD'	DTA'	DIQ	DKR	DLS	DWU
	DPX	DOV	SE'F'	DD'F'	DWE'	DYC'	DZD'	EIR	EPD'	EGT	EMZ	EOX	ENF'	EKQ	EPV	FGW
	FJZ	FIU	VA'F'	FKF'	EUD'	EWC'	MA'C'	EYB'	GC'D'	GYZ	GXF'	RUD'	LB'D'	GUV	GRS	GPD'
	GNO	SWD'	GRB'	GKL	GJA'	RVC'	FTC'	FYA'	FLX	FNQ	SVD'	FRD'	WZA'	FPS	RXB'	OA'D'
	SUC'	IJK	ICB'	INW	ILA'	IPC'	IST	IVW	IZP'	IXY	A'BD	ABD	AES	AC'D'	AXZ	ANE'
	AFE'	AGI	AJL	AKA'	AEO	BCI	AUW	ABT	ACQ	AYF'	APV					

ABCF	QTA'	TB'F'	BEA'	BDG	OFX	BFB'	NQU	PQX	MNZ	PRV	DPF'	BOU	BXD'	BKS	BJT	MSY
	ELB	BNV	OQV	BAW	CEL	BYD'	BZC'	CMV	QWB'	CKT	CPO	MTX	CGC	CJS	MPA'	LVF'
(I)	LWZ	PSU	UYA'	LUX	LQS	A'BD	NTY	JOC'	OSF'	JMD'	JND'	JPB'	TWD'	JWX	OTZ	JVY
	QYC'	JUZ	KED'	KCF	QZE'	KMD'	KQR	KWY	TVD'	KVX	KUF'	NSX	KB'C'	NRF'	LOD'	LMC'
	LEP	TUC'	CRA'	CUB'	CPZ	CNW	DFN	CC'F'	CD'F'	CYD'	CXD'	DEK	DSA'	DVB'	DOW	DPY
	NA'C'	DLT	DMU	DJR	DEQ	DD'F'	DZD'	DXC'	VZA'	EFD'	EGS	EHR	ENZ	EMF'	EOY	EJQ
	SZB'	EPW	FEU	FGV	FJF'	FKZ	EUD'	EVC'	EXB'	GYF'	GXZ	GC'D'	RUD'	GHT	GPD'	GOW
	GNB'	SWD'	GKA'	GMO	GJL	FXA'	FLY	SVD'	FMQ	FRD'	FSC'	FPT	HLA'	HMN	LB'D'	HJK
	RWC'	HOB'	CA'D'	HPC'	HST	HXY	HZF'	HVW	RYB'	HD'D'	WA'F'	AXF'	AYZ	AUV	AC'D'	AFW
	AMB'	APE'	AIQ	AET	AJA'	BCH	ARS	ANO	AKL	AGH	ACD					

ABDE	BNY	BMX	ECA'	BOZ	BVE'	BUC'	QUD'	UVA'	RD'F'	BWD'	YZA'	OQY	BXY	HQF'	MRU	PQW
	TZC'	NQZ	CFW	BDL	BFF	BHS	ASV	BGR	BB'F'	QVC'	MTW	CEG	CDA'	CMY	PRY	CLC
(J)	XA'F'	CFC'	CIS	CHT	LUW	LPD'	TYD'	LRT	CXB'	TXD'	PTF'	NTV	OTU	KOB'	KMN	KLA'
	KPC'	LXZ	KST	KXY	KZF'	NSW	KVW	LMO	KD'D'	PSZ	LNB'	CPU	COF'	CNX	DFD'	DGT
	CWE'	CVD'	DIH	CZB'	DUD'	DOX	DWF'	DPV	SC'F'	RSA'	DHZ	DKQ	DIR	EFR	DWC'	DYB'
	EIQ	EOV	ENU	EKR	ELS	EPX	PA'D'	ED'F'	SYD'	FHZ	FGY	FIF'	NA'B'	FKU	EZD'	EYC'
	ETA'	EWB'	GVF'	GWZ	SXE'	HIK	GB'D'	GQS	GUX	GMC'	LC'D'	GNP	GOD'	RVB'	GIL	GHA'
	FIE'	FNA'	FLV	FNR	FCS	FQD'	HMD'	HOP	HND'	HQR	RXC'	NOA'	HNY	HOF'	HVX	SUE'
	RZD'	IUZ	A'CD	IVY	IWX	IND'	IMD'	HB'C'	IPB'	IOC'	LYF'	ADE	AWF'	AXA'	APX	ANC'
	AOE'	AQT	ACK	AHL	AMF	BCK	AB'D'	AVZ	AUY	AES	ACR					

ACD	WC'D'	LPQ	LMT	YB'D'	LCB	LNS	KOS	KNR	KTB'	KQD'	QA'C'	KWA'	MVB'	JQD'	JSB'	JOT
	JLW	JMR	KLV	JVA'	NPX	QTY	OPF'	BA'E'	EGM	BEY	BDW	CKC'	CLE'	CIB'	CHO	MPY
(U)	CJP	XB'D'	CEX	CGN	CEV	BA'D'	DJC'	BKP	EIO	EHB'	BSP'	BTZ	BRX	BQV	MSA'	BLD'
	IPS	IRE'	B'CF	ILX	INQ	IYA'	NTA'	ITC'	IJZ	IKP'	HXA'	PVD'	MPT	HSC'	HRD'	NOV
	HLY	HMQ	HJF'	HKZ	FZF'	FVW	FXY	FLA'	FMN	NWB'	FPC'	WYF'	FSZ	OZC'	PWD'	FOB'
	GHV	GIW	FD'D'	GRC'	GCE'	GSD'	WXZ	GTD'	GKY	GJX	GLZ	GA'F'	ND'F'	PZB'	VC'D'	OYE'
	OXE'	FJK	MZE'	EQZ	ETV	ESW	FHI	EOA'	ELB'	WYC'	EID'	EJN	EKM	EGP	ETW	EHD'
	NZD'	EFR	DHN	DIN	DKD'	DLC'	DJD'	MXC'	VYZ	DPA'	MOW	VXP'	DSY	DTX	DRZ	BSV
	CA'E'	CEF'	CFF'	DGO	DFQ	QSX	CQW	CTF'	CSZ	ND'F'	CRY	ASD'	AOQ	ATD'	AZA'	BCF
	AEE'	ABN	ACM	APR	ALF'	AKX	AHW	AJY	AIV	AEC'	AFG					

ACE	MBA'	OPX	QTZ	NPF'	QSF'	WD'D'	BHN	BGB'	BEZ	CLC'	CJD'	CKD'	CIN	MPZ	CHN	CFQ
	CIF'	CGO	XE'C'	BA'C'	CDU	BIN	BJD'	BSX	BTY	STW	BQU	BRF'	BKD'	BLP	LQD'	LTB'
(V)	UYZ	UXF'	LWA'	LOS	LNR	KMT	KOR	OTA'	KPQ	ZE'D'	KNS	QA'E'	UC'E'	JUA'	JRB'	NUB'
	JQC'	JMS	JKN	KLU	JKT	ISD'	OD'F'	IPR	ILF'	B'DF	PUD'	IOQ	IZA'	ITD'	IJY	IEX
	RA'F'	HTD'	WZF'	HQB'	HEC'	HSD'	HLZ	HJX	MKY	HIW	FXZ	FYF'	FUN	FHO	FNB'	PWC'
	OZD'	FPD'	FRT	FC'D'	GHU	GRD'	GSC'	WXY	GLY	GKZ	GJF'	HC'F'	PYB'	GXA'	NOU	GHQ
	GFT	OYC'	FJL	FKA'	ECY	ETU	ZRW	MNW	NXD'	FGI	SA'B'	ENA'	NYD'	NYD'	EJO	EIC'
	ELM	EKB'	FGE'	EHP	NZC'	DA'D'	OWB'	EPS	EJP	DHO	DIB'	DLD'	QRX	DKC'	DQW	NXD'
	DRY	DSZ	DTF'	RSU	DEX	DGN	CSY	CTX	CRZ	CPA'	BDF	APS	AJZ	AIU	AZD'	BCL
	AFE'	ABO	AIM	AYA'	APH	ANQ	AKF'	ALX	AGW	ATC'	ACB'					

ACF	BLC'	BKD'	NFA'	BJD'	BIN	BTX	BSY	NPF'	STV	BPA'	BRZ	QSZ	NPZ	BEF'	BDU	BHE
	BGO	EFQ	OFY	QTF'	CLP	CJD'	CKD'	CNM	CGB'	UXZ	CIN	CEZ	CDF	TA'B'	YB'C'	LQD'
(W)	LSE'	LVA'	OSA'	KQC'	KBE'	KNT	ZB'D'	KMS	QA'D'	UC'D'	KUA'	LOT	LMR	JLU	ZA'F'	JKV
	JMT	JOR	UYF'	JNS	JPQ	OD'F'	IQB'	ISD'	IRC'	PUD'	ILZ	ITD'	IJX	B'DF	IKY	HTE'
	PVC'	HZA'	NUB'	HSD'	HFR	HLP'	HOQ	HKX	HIV	HJY	OZD'	FYZ	FXF'	PUV	FNO	MNV
	FMB'	FPD'	FRS	FC'D'	PXB'	GIU	GRD'	GTC'	GLX	GKF'	GJZ	VD'D'	GYA'	GNQ	GPS	FJA'
	FKL	OXC'	ECX	ESU	ERV	NXD'	FGH	MZC'	EMA'	VZP'	EKO	NYD'	EJB'	ELN	EGD'	EHC'
	EIP	NYD'	EFT	DA'D'	RTU	OVB'	QRY	DIO	DHB'	DLD'	DJC'	DKP	DQV	DRX	MOU	DSF'
	DTZ	MXD'	CA'C'	VXY	DGN	DEY	NC'F'	CSX	CRF'	CTY	CQU	ADN	ACO	AMQ	AJF'	AGV
	AKZ	ALY	AHU	ECV	AXA'	AFI	AED'	APT	ASC'	ABD'	ABB'					

ADE	STY	OPV	RA'E'	BEF	BEZ	BGC'	MQA'	UWZ	BLO	UVF'	BID'	BJM	BHD'	BKN	BSV	BTW
	YZF'	BRU	EQF'	CLB'	CKB	CHD'	CID'	CJN	CGP	CFR	CEU	CDF'	DA'B'	LRD'	LTC'	KSD'
(X)	KLF'	KOQ	KFR	UB'D'	KZA'	KTD'	YD'D'	VWY	LNQ	NVD'	LPS	JA'F'	PTA'	ZC'D'	MUC'	JGB'
	JTE'	JSE'	JBC'	JKY	LYA'	JLZ	IPQ	IMT	NOF'	IOR	INS	ILU	IJW	C'DF	IKV	NPU
	HEB'	HQC'	HRS	HNT	HJV	PWB'	HKW	HIY	HUA'	FWF'	FUY	FVZ	FMP	FNC'	MB'F'	FGD'
	NWE'	FQT	GHF'	GIZ	FE'D'	GSB'	NVD'	GQD'	GJU	GLW	PYC'	GVA'	GMR	GOT	EA'D'	ETF'
	PZD'	OYB'	VE'C'	FIA'	ESZ	EBY	EQW	FHL	MNY	FGK	ELD'	RTX	EIB'	EJP	EKC'	EHO
	EGN	SA'C'	NZB'	OWC'	RSF'	PD'F'	DHP	DJO	DIC'	MOZ	DKB'	DLM	QRV	DNA'	DQY	DIU
	ERW	EWD'	EGE'	DEV	DFS	CQZ	OUD'	CTV	CSW	QSU	COA'	AEM	ABP	BCY	AWA'	AQD'
	ANJ	ALV	AOS	ANR	AIF'	AHZ	ADD'	ACC'	APJ	AGY	ATB'					

ADF	XD'D'	BLB'	EID'	EHD'	BKM	BJN	BTV	BSW	STX	BOA'	UVZ	BQZ	OPW	XZF'	RA'D'	BEU
	BIF'	BFR	EGP	MQA'	CLO	UWF'	CKN	CHD'	CJM	CID'	CGC'	CEF	CDZ	TA'C'	LRD'	LSC'
(Y)	QTU	KRC'	UE'D'	KQE'	KSE'	KLZ	LXA'	KA'F'	KTD'	LPT	VWX	PSA'	LHQ	NVD'	JOQ	JPR
	JZA'	ZC'D'	JSE'	JTD'	JKX	JLF'	IRB'	WB'C'	IQC'	IMS	INT	IKW	IJV	IUA'	HPQ	HOE
	HMS	HRT	HLU	HJW	HKV	HIX	PVB'	C'DF	MUC'	NOZ	FVF'	FWZ	FUX	FMC'	FNP	NWE'
	FOD'	FQS	GHZ	GIF'	FE'D'	GTB'	GQD'	NVD'	GKU	PXC'	GLV	GOS	GNR	GWA'	EA'D'	EZY
	PZE'	ESF'	ERX	EQV	MZB'	OXB'	MNX	FGJ	ETP'	FIL	FHA'	ELD'	EJC'	EKP	EHB'	EIO
	EGM	PD'F'	EMC'	DIP	DJE'	NSZ	DKO	DMA'	ELN	QRW	OVC'	DQX	DRV	DSU	NB'F'	MWE'
	CA'B'	DGD'	DFT	MOF'	DEW	CQF'	CRU	CSV	CTW	MPU	OUD'	BCX	AOT	AQD'	ALW	AJU
	AFK	AHF'	AIZ	AMR	AVA'	AEN	ADD'	ACP	ABC'	ASB'	AGX					

AF	OCS	NQR	XZC'	QTB'	PRS	MEB'	BDC'	BGZ	BHY	BFA'	TA'F'	XYD'	CDP	DWC'	CGF'	MFB'
	CHX	CIY	CJV	CKW	QWA'	CMS	CPL	CEO	BPQ	BOR	BJW	BIX	BKV	BMT	BNS	BLU
(E')	ITZ	LSF'	LRX	UYB'	LCV	JKD'	JLC'	JTX	JSY	JRZ	KLP	HUZ	JPA'	ZD'F'	PUV	YC'F'
	OUX	KA'C'	KTY	KSX	KEF'	KQU	MVX	CQC'	CRB'	CUA'	CWT	MKZ	DGX	VKD'	DFK	DEN
	DSB'	DVA'	DCT	DQD'	OWZ	OVF'	DLW	NB'C'	DNE	DJU	DHF'	DIX	NOC'	EFI	VZB'	PC'D'
	EGV	EHU	EJF'	EKZ	ELY	EMQ	FHZ	FGS	SZA'	ESC'	ETC'	ERD'	EPT	END'	FJQ	EXA'
	NWY	PYZ	GNA'	PXF'	GJO	GLM	GKB'	GQY	NOP	GTU	GRW	GHP	GIC'	FPW	FMF'	FNZ
	FCY	FXE'	FUD'	FVC'	OB'D'	HTV	HJN	HKM	HID'	RLB'	LA'D'	NVX	HOA'	HSW	NUF'	HQZ
	IA'B'	ITW	WE'F'	ILO	IJM	RYA'	IKN	IQF'	IRU	STD'	ISV	ALN	AKO	AMA'	AGD'	BCE'
	AEW	ACZ	AFV	ASU	AFT	AQX	AJB'	AHC'	AIF	ADY	ABF'					

ABC	MBD'	BGH	BFW	BET	CMO	CKA'	CJL	CGI	CFV	CES	BC'D'	BYZ	BXF'	BUV	MSC'	BPD'
	BRS	BMB'	BJA'	BKL	LVD'	LWD'	LPU	NRD'	LOF'	LNX	LMY	KUD'	KYB'	KWC'	KPV	KNF'
(Q)	KMZ	KOX	JXE'	JVC'	JUD'	JMF'	JPW	JOY	JNZ	ID'F'	IWB'	NPS	ITA'	IZD'	NTC'	IYC'
	IPX	INU	ILS	LZB'	ICV	HD'F'	IKR	HXC'	HVB'	HZD'	HPY	HSA'	HLT	HMU	HOW	HJR
	GC'F'	GYD'	GXD'	GRA'	GUE'	GMV	GNW	GPZ	GKT	GJS	FPA'	FSY	FRZ	FLC'	FIX	EWX
	EVY	EUZ	FKE'	FJD'	FGO	FIM	FHN	EPB'	EMD'	NYA'	EHD'	EOC'	EGL	EFF'	EHK	EIJ
	DBY	DZF'	DC'D'	DVW	DST	DPC'	DOB'	MXA'	DMN	DLA'	DJK	CXZ	CYF'	DEB	CC'D'	DFU
	DHI	CUW	MFT	CRT	CXB'	OPR	CPD'	TVZ	OTD'	UA'C'	OSD'	TWF'	A'BF	TB'D'	RB'C'	SUX
	WA'E'	RUF'	RVX	OZA'	RKY	SVF'	SWZ	VA'E'	SB'D'	TUY	BCD	AZC'	AYD'	AXA'	AJT	AMX
	AKS	ABI	ACH	ADG	AFE'	AOU	AXD'	ANV	ALR	AFB'	BNO					

ABD	LFF'	YA'D'	CE'C'	LNV	LCU	LMW	KD'F'	KZD'	MVA'	KYC'	KWB'	PSD'	KPX	KTA'	KNU	KOV
	UA'B'	PTE'	JSA'	LXD'	JZD'	JVB'	JXC'	KLS	JD'F'	JPY	JOW	TXZ	JLT	JMU	MQD'	TYF'
(R)	BGJ	EFY	XA'E'	TC'D'	BET	CHL	QWY	CIA'	CGK	NQD'	CFX	CMP	BB'E'	CDS	BMC'	MSB'
	QVX	BHA'	BIL	QUF'	BWZ	BUX	RVF'	BQS	BOD'	IPV	IKQ	SUV	IMZ	IOX	INF'	LZC'
	IJD'	A'CF	IYB'	IWC'	LYD'	NTB'	HVC'	HUD'	HXB'	HPW	HJQ	HOY	HNZ	HMF'	FOA'	FSW
	FQZ	FLB'	FTV	GIT	GHS	NWA'	GNY	GOZ	GMX	GQA'	GB'F'	GUD'	GVD'	GUC'	NOS	PZA'
	EXY	EVW	SXF'	SYZ	EZF'	FKM	FID'	FJN	FGP	FHD'	ED'D'	EPC'	EST	EJX	EOB'	EMK
	ELA'	EHI	SC'E'	MOT	EFU	DHK	DIJ	DMD'	DPB'	DOC'	DND'	DVY	DWX	TOW	DUZ	DEQ
	CB'I'	CJF'	DFF'	DGL	CCT	CNC'	COD'	CVZ	CUY	OPQ	APU	ACJ	ABK	ANX	AFC'	ALQ
	ABT	AOF'	AZE'	BCZ	AED'	AEG	ADA'	AVD'	AIS	AMY	BMP					

ABE	BGA'	BHJ	EFZ	XA'C'	QUX	MQC'	TD'D'	TZF'	MRB'	QWZ	CHK	CIJ	QVF'	CGL	CEQ	CFF'
	RC'D'	CDR	PFD'	BB'C'	BIK	BWY	BVX	BUF'	CQD'	BQE	BOP	BND'	BMD'	LPX	LWB'	LTA'
(S)	LNU	TVW	LOV	QE'D'	KMW	KOU	ZA'D'	KNV	OTB'	KPF'	KXD'	KYD'	KZC'	JPZ	TXY	JNW
	JMV	LYC'	JKT	KLR	PTC'	JC'F'	JYD'	JXD'	JUB'	MUA'	JRA'	NPQ	IPU	IOF'	IMY	INX
	LZD'	A'DF	IVD'	IWD'	IZB'	HB'F'	RYZ	ILQ	HUC'	HVD'	RXF'	HWD'	HOZ	NQA'	HNY	HMX
	HIT	FNA'	FEW	FQY	MMT	FTU	GHR	LD'F'	GJQ	GPW	GNZ	GMP'	GOY	PYA'	NOR	GXB'
	GUD'	GVC'	EYF'	EXZ	EUW	FIC'	RUV	FJO	FLM	FKB'	EC'D'	FHP	FGD'	EPD'	ERI	ENE'
	EMO	EKA'	VA'B'	EJL	EGI	EFV	DWF'	DB'D'	OWA'	DHL	DIA'	DMP	DOD'	DNC'	DQT	DVZ
	EUY	CWX	DGK	DPX	CND'	COC'	CPE'	CMD'	CUZ	CVY	ADJ	AGT	BDE	BCT	AFD'	ACX
	AIR	ABL	ANF'	APV	AWC'	AYB'	AUD'	AMZ	AKQ	ACA'	AEH					

ABF	LVE'	QWF'	YA'C'	LPY	OSB'	LSA'	QB'D'	LOW	LMU	KC'F'	ZA'D'	KNW	KMV	KPZ	KRA'	EUB'
	MTD'	KXD'	FSC'	JPF'	JCU	JNV	IB'F'	JMW	JKS	JLR	JZC'	JXD'	JYD'	LXC'	BFF'	BGL
(T)	QUY	BHK	EEQ	EDR	NCC'	CHJ	CIK	CGA'	CFZ	PRD'	CDE	NRB'	BIJ	BUZ	OQD'	BVY
	EWX	BPB'	BOC'	BMD'	QVZ	BND'	BC'D'	IQA'	RYF'	A'DF	IMX	INY	IOZ	IVD'	IUC'	IWD'
	WA'B'	NOA'	IZD'	HWD'	HVD'	HZB'	HOF'	HPU	HMY	SVW	HNX	RXZ	MLQ	HIS	FMA'	FRV
	FQX	MSS	FSU	GIR	PXA'	GKQ	LD'F'	GPV	SXY	GHZ	GOX	GHP'	GUD'	GWC'	GYB'	EYZ
	EXF'	EUV	FKO	FJB'	FIP	RUW	FLW	FHC'	EC'D'	FGD'	ERS	EPD'	ENO	SZF'	EHB'	EJA'
	EKL	EGN	MCE	EFW	DWZ	DB'D'	DIL	DHA'	DMC'	SD'D'	DNP	DQS	DOD'	DVF'	DUX	CWY
	OVA'	CB'C'	DFY	DGJ	MPQ	CWD'	COP	CQR	CMD'	CVX	CUF'	APW	AGS	AMP'	ACL	AEI
	AFD'	AOY	AHS	AUD'	BCS	AXB'	AVC'	ANZ	ADK	ABA'	AJQ					

A	LTE'	LSD'	LPR	LOQ	KYA'	KTC'	KBD'	KPS	KNQ	KLX	JXA'	JSC'	JRD'	JPT	JMQ	JLY
(F')	JKZ	IWA'	ITB'	IQD'	ICS	LZA'	INR	ILV	IKU	HVA'	HSR'	HQD'	HOT	HMB	HLW	HJU
	HIZ	GUA'	GRE'	GQC'	GNT	GNS	GKW	GJV	GIY	GHX	FWX	FVY	FUZ	FPB'	FOC'	FND'
	FMD'	FIJ	FHK	FGL	ETX	MNZ	ESY	ERZ	EPA'	ELC'	EKD'	EJD'	BIM	EHN	EGO	EFQ
	DTV	DSW	DCZ	DOA'	DLB'	MOY	DKM	DJN	DID'	DHD'	DGP	DFR	DEU	CTU	CRW	CQY
	CNA'	CLM	CKE'	MPW	CJO	CIC'	CHP	CGD'	CFS	CEV	CDX	BSU	BRV	BQX	BMA'	BLN
	BKO	BJB'	BIP	BHC'	BGD'	BFT	BEW	BDY	QTW	QA'B'	HVC'	YC'D'	VWZ	UB'C'	ZD'D'	MUC'
	TA'E'	XYZ	OFU	NPV	UVX	RA'C'	UWY	XC'D'	STZ	WB'D'	WOX	NUD'	OZB'	NWC'	PXD'	PYD'
	PZC'	VB'D'	BTY	NYB'	SA'D'	OWD'	OVD'	QBU	MXB'	ESX	QSV	AOB	AHY	AJW	AKV	BCZ
	ANS	ACD'	AEC'	ABE'	AMT	ALU	AIX	AGZ	AFA'	AEB'	APQ					

AB	BBI	BDK	BGS	MSU	BHR	BFD'	MQX	EJQ	BMF'	BNZ	BPW	BOY	BXB'	BVC'	BUD'	CDJ
(A')	CEN	CGT	CFD'	CIE	CEZ	CKQ	LST	LPC'	LVW	MRV	JKL	IB'D'	JPE'	JNO	JNB'	JUV
	JES	JYZ	JXB'	JC'D'	KEB'	KMO	LXY	KRT	KPD'	KYF'	KXZ	KUW	LOB'	LSN	KC'D'	CNF'
	CPV	COX	CUE'	DFC'	DEG	CYB'	CWC'	DIS	EHT	DMY	DLQ	DPU	DOF'	DNX	DVD'	DZB'
	DWD'	EPB'	EKS	ELR	EJT	EOU	ENV	EMW	EPF'	FHY	FGZ	FKV	FJW	FIX	EZC'	EYE'
	EXE'	FNS	FOR	FKT	FLU	FPQ	GHJ	GIK	LE'D'	GMD'	GOP	GND'	GQR	HIL	GB'C'	GWY
	GVX	GUF'	HNF	HMC'	HCS	HOD'	LZF'	HWZ	HVF'	MUX	HB'D'	INC'	IMP	IQT	IOD'	IUY
	IWF'	IVZ	CB'F'	SD'F'	SZD'	RUB'	SXC'	SVB'	RXD'	RYD'	QVD'	PEZ	RC'F'	OQZ	NRW	QUC'
	TE'F'	NQY	TZE'	PTX	TWB'	NTU	OTV	PSY	CSW	TYC'	QWD'	ECL	AVY	AMD'	AOC'	AWX
	ACS	ASQ	ADR	ABT	AUZ	AND'	AHK	AFF'	AGL	APB'	AIJ					

AC	QTE'	CSD'	WE'F'	MRS	BDN	BFI	BZD'	BHU	BGV	BLY	BKZ	BMQ	BPT	BRD'	BXA'	NBT
(B')	BSC'	BJF'	CEM	MPD'	CGW	UXD'	CED'	CFH	OST	CIU	CLX	CKF'	CJZ	NPD'	OPC'	LSB
	LTV	LQZ	KIM	JLN	IA'E'	MUV	JKO	UYD'	JMA'	JRV	JQX	JSU	UC'F'	QA'F'	KRW	KQY
	KTU	TWA'	KNA'	LCA'	CRE'	CTC'	CNQ	CPS	OVW	DPG	DEC'	WC'D'	CYA'	QRC'	DTD'	DSD'
	DPE	DOQ	ELF'	DKI	VXC'	DHW	DIV	DJY	MYZ	EPA'	MXF'	DZA'	EHY	NYF'	EGZ	EKV
	EMW	EIX	EMT	OXY	ENS	ELU	VZD'	EOR	EPQ	FJT	FKS	RUA'	NXZ	GA'C'	VD'F'	GJD'
	GLF	GKD'	GTY	GSX	GQU	GRF'	MC'D'	GHM	MNO	GIN	FPF'	FMW	FNV	FOU	FLR	WYC'
	FYD'	FZC'	FXD'	EIO	HKP	OZF'	HJC'	SVA'	HLD'	PWX	HQV	WZD'	HRX	HTZ	HSF'	HA'D'
	FVY	IKC'	IJP	ITF'	PUZ	NUW	ILD'	ISZ	IBY	IQW	OD'D'	ABZ	ALC'	AFQ	APA'	AIH
	ATX	ASY	ADU	ADW	BCO	AJD'	AGO	AZF'	AKD'	AMW	ACV					

AD	TYA'	OPB'	LTX	LPZ	LSY	LPA'	PST	MUX	JQV	JLD'	JKP	JSF'	JTZ	JRX	JA'D'	HVF'
(C')	OUZ	KTF'	KA'D'	KLD'	KEY	KQW	KSZ	UB'F'	YE'F'	BFK	BCX	BED'	RA'F'	BEN	XZD'	KQZ
	CEM	CDD'	CSJ	CGY	CIF'	CHZ	UWD'	CKU	CLV	QUA'	MQS	BQD'	BOT	BVA'	BSB'	YZD'
	EHF'	BJU	UVE'	DIX	XD'F'	BLW	BNR	COS	MWZ	CWB	CTB'	CQD'	DEB'	MOD'	DFA'	DGZ
	CVY	QRB'	NB'D'	CWA'	VXE'	DNS	RSD'	DPQ	OWX	DOB	DKV	DLU	DMX	DIX	DJW	DHY
	EFG	PE'D'	EFW	EJY	EKX	EIV	EOQ	ELF'	RTD'	FHT	ESD'	EPR	PZF'	FIS	ETD'	EZA'
	GA'E'	SXA'	GJM	GLO	GKN	GRU	GQF'	GTW	GSV	GID'	GHD'	FPU	FOF'	FWY	MB'D'	FNX
	PXY	FIQ	WYB'	NWF'	FZE'	FWD'	MNP	FVD'	HKO	HJB'	HIP	HLN	NVZ	HMA'	HSU	HQX
	HEV	ILM	IKB'	NOD'	IJO	INA'	NUY	PVW	IEW	ITU	IQY	AQZ	AID'	AJN	ASW	AEU
	ACX	ADY'	AEY	ATV	BCP	AND'	AGP	AFR	ALB'	AKM	AOA'					

AE	PRT	OQT	TZA'	XYD'	QSE'	BFL	BGF'	BHX	BEO	BDP	CDC'	XC'F'	QVA'	CHY	CIX	CJW
(D')	CIU	CMT	CKV	CGZ	CEB'	CFA'	UXB'	BQC'	ERB'	MQR	YZC'	BUA'	BJV	UVC'	BKW	BIY
	BMS	LSZ	LEY	LCW	LTF'	NPB'	JKD'	PUW	JLP	JTY	JSX	JRF'	JA'C'	ZD'F'	JQU	OUY
	KIC'	MUF'	NVX	KTX	KEZ	KSY	KPA'	CNS	COR	MWY	CPQ	DFJ	DGY	VWD'	OVZ	DEM
	DQE'	DTB'	DNR	DOS	ESC'	DKU	DLV	DHZ	DIF'	EFH	DWA'	OWF'	SA'F'	EGW	PC'D'	MCP
	EJZ	EIU	EKF'	MNE'	ELX	ENQ	FGT	EPS	ERD'	FIR	FKQ	ETC'	VB'F'	EYA'	PYF'	NWX
	GNA'	GJB'	GKO	GLN	GFV	GSU	GQX	GHC'	GIP	FPV	WZB'	FMZ	FOX	MB'C'	FNF'	PXZ
	FWC'	FUD'	FYB'	HTW	NCC'	HLO	HJM	HIE'	HKN	LA'D'	RXA'	NVY	HRU	HQF'	HSV	IJN
	IKM	NUZ	HA'B'	OB'D'	ICA'	ILB'	IQZ	ITV	ISW	STD'	BCD'	AKB'	AGD'	AHP	ANA'	ADX
	ACF'	ABZ	AQY	AIC'	AFS	ATU	ARW	AJO	AEV	ALM	BNT					

APPENDIX III-C

TWO-FACTOR INTERACTIONS ALIASED IN STRINGS

F		EF		E		DF		DEF		DE	
AB	NU	AG	MN	AC	MU	AH	MO	AQ	LR	AI	MB'
CG	OV	BC	OB'	BG	OW	BD	NB'	BI	MW	BQ	NO
DH	PX	DQ	PC'	DI	PY	CQ	PD'	CH	NV	CD	PD'
EJ	TA'	ER	ST	EK	SA'	ES	RT	DG	OU	ET	RS
FM	WB'	FU	VW	FN	VB'	FV	UW	EA'	PF'	FW	UV
IQ	YC'	HI	XY	HQ	XC'	GI	XZ	FB'	XD'	GH	XF'
KR	ZD'	JK	ZF'	JR	ZD'	JL	YF'	JT	YD'	JA'	YZ
LS	D'F'	LA'	D'D'	LT	D'F'	KA'	C'D'	KS	ZC'	KL	C'D'

BCE		BCDF		BCDEF		BCDE		BCD	
AY	BC'	AD'	BZ	AF'	BD'	AD'	BF'	AZ	BD'
CP	DD'	CF'	DX	CD'	DC'	CZ	DY	CD'	DP
EN	FK	EV	FS	EB'	FA'	EW	FT	EO	FL
GX	HF'	GD'	HP	GZ	HY	GD'	HC'	GF'	HX
IZ	JU	IC'	JO	IX	JW	IP	JB'	IY	JV
LW	MR	KB'	LM	KV	LU	KO	LN	KW	MS
OT	QD'	NA'	QY	MT	NS	MA'	QX	NT	QC'
SB'	VA'	RW	TU	OR	PQ	RV	SU	RB'	UA'

D		CF		CEF		CE		CDF		CDEF	
AD	MV	AJ	MP	AR	LQ	AK	MC'	AS	KQ	AA'	IJ
BH	NW	BE	NC'	BK	MY	BR	NP	BL	MZ	BT	MD'
CI	PZ	CR	OD'	CJ	NX	CE	OD'	CA'	NF'	CS	ND'
EL	RA'	DS	QT	DA'	OF'	DT	QT	DJ	OX	DR	OC'
FO	UB'	FX	UV	EG	PU	FY	UY	EH	PV	EQ	PB'
GQ	XD'	GK	VZ	FC'	VD'	GJ	VZ	FD'	UD'	FF'	UZ
JS	YD'	HL	WF'	HT	WD'	HA'	WF'	GT	WC'	GL	VY
KT	C'F'	IA'	B'D'	IS	ZB'	IL	B'D'	IR	YB'	HK	WX

BC		B	
AF	BX	AF	BM
CY	DZ	CN	DO
EP	GC'	EP	GU
HD'	ID'	HV	A'F'
JM	KN	IW	JX
LO	QF'	KY	LZ
RU	SV	QB'	RC'
TW	A'B'	SD'	TD'

CDE		CD		C		BF		BEF		BE	
AT	JQ	AL	MD'	AE	MX	AM	BF	AU	BN	AN	A'D'
BA'	MF'	BS	ND'	BJ	NY	CU	DV	CM	DB'	BU	CF
CL	NZ	CT	OP	CK	OZ	EX	GN	EC'	FG	DW	EY
DK	OY	DE	QR	DL	QA'	HO	IB'	HW	IV	GK	HB'
EI	PW	FZ	UF'	FP	UC'	JP	KC'	JY	KX	IO	JC'
FD'	UD'	GA'	VX	GR	VD'	LD'	TF'	LF'	TD'	KP	LD'
GS	VC'	HJ	WY	HS	WD'	QW	A'D'	OQ	ZA'	LC'	PA'
HR	XB'	IK	B'C'	IT	B'F'	RY	SZ	PR	SD'	RZ	SY

LEGEND

Original Factor Label → F

Strings of Aliased Two-factor Interactions:

F	
AB	NU
CG	OV
DH	PX
EJ	TA'
FM	WB'
IQ	YC'
KR	ZD'
LS	D'F'

BDF		BDEF		BDE		BD		BCF		BCEF	
AV	YA'	AB'	BW	AW	BB'	AO	BV	AX	BP	AC'	BY
BO	CB'	CV	DU	CO	DN	CW	DF	CC'	EM	CX	DF'
DM	ED'	EF'	FQ	ED'	FI	EZ	GB'	FJ	GY	EU	FR
FH	GW	GO	HN	GV	HU	HM	IN	HZ	IF'	GP	HD'
IU	JZ	IM	JD'	JF'	KZ	JD'	KD'	KU	LV	ID'	JN
KF'	LX	KD'	LC'	LY	MQ	LP	QU	NR	OS	KM	LB'
NQ	PS	PA'	RZ	PT	RD'	RF'	SX	QD'	TB'	OA'	QZ
RD'	TC'	SY	TX	SC'	XA'	TY	A'C'	DD'	WA'	SW	TV

2_{IV}^{32-26} APPENDIX III-D. INNER-PRODUCT SUMS* N = 64

Trends		A	B	C	D	E	F	G	H	I	J	K	L	M	O	P
Linear	LX															
Quadratic	QX															
Cubic	KX															
Original Factorial Labels		ABCDEF	ABCDE	ABCDF	ABCEF	ABDEF	ACDEF	ABCD	ABCE	ABCF	ABDE	ABDF	ABEF	ACDF	ACEF	ADEF

	Q	R	S	T	U	V	W	X	Y	Z	A'	B'	C'	D'	E'	F'
LX																64
QX											64	128	256	512	1024	
KX	1280	2560	5120	10240	5120	10240	20480	20480	40960	81920						43648
	ABC	ABD	ABE	ABF	ACD	ACE	ACF	ADE	ADF	AEF	AB	AC	AD	AE	AF	A

	AF'	AF'	AE'	AD'	AC'	AB'	AA'	AZ	AY	AX	AW	AV	AU	AT	AS	AR	AQ
LX																	
QX																	
KX								10240	20480	40960	40960	81920	163840	81920	163840	327680	655360
		BCDEF	BCDE	BCDF	BCEF	BDEF	CDEF	BCD	BCE	BCF	BDE	BDF	BEF	CDE	CDF	CEF	DEF

	AP	AO	AN	AM	AL	AK	AJ	AI	AH	AG	AF	AE	AD	AC	AB
LX											128	256	512	1024	2048
QX	256	512	1024	2048	1024	2048	4096	4096	8192	16384					
KX											86976	171392	322304	480768	-349184
	BC	BD	BE	BF	CD	CE	CF	DE	DF	EF	B	C	D	E	F

---

*See footnotes in Appendix I-C.

APPENDIX IV

COMPUTER PROGRAM FOR OBTAINING SCREENING DESIGN ALIASES

```
STMT /*                             PROGRAM ALIAS                                 */
     /*     WRITTEN BY HOWARD B. LEE                                              */
     /* THIS COMPUTER PROGRAM HAS BEEN WRITTEN IN PL/1 FOR AN IDM 360/91          */
     /* COMPUTER.                                                                */
     /* PROGRAM FOR COMPUTING TWO AND THREE WAY ALIASES FOR FRACTIONAL            */
     /* FACTORIAL DESIGNS (SCREENING DESIGNS) .                                   */
  1  ALIC:PROC OPTIONS(MAIN);
  2   DCL GOOG FILE STREAM OUTPUT;
  3   DCL P(32,8) CHAR(1),BZ(8,32) CHAR(1),M(32) FIXED BINARY(15,0);
  4   DCL ZZ CHAR(6),A(32,8) CHAR(1),MP(32) FIXED BINARY(15,0);
  5   DCL MM(32) FIXED BINARY(15,0),PE CHAR(2);
  6    DCL R(4961) CHAR(8),Q(4961) CHAR(8);
  7   DCL LCC(4961) FIXED BINARY(15,0);
  8   DCL  KCC(4961) FIXED BINARY(8,0);
  9    KIP=15;
     /*                                                                           */
     /* THIS ROUTINE IS USED TO COPUTE THE ALIASES FOR BOTH TWO FACTOR            */
     /* INTERACTIONS AND THREE FACTOR INTERACTIONS.         */
     /*                                                                           */
 10   ALIAS:PROC(N,NE,M,MP,X,L,LL,MZ,P,KP,KB,TRIP);
 11    DCL S(32,8) CHAR(1);
 12   DCL X(*,*) CHAR(1),P(*,*) CHAR(1);
 13   DCL (M(*),MP(*)) FIXED BINARY(15,0);
 14  DCL (N,NE,L,LL,KP) FIXED BINARY(15,0);
 15   DCL TE CHAR(1);
 16    DCL ZE CHAR(1);
     /*                                                                           */
      /*        COMPUTATIONS TO FIND THE TWO FACTOR INTERACTION TERMS             */
     /* CHECKS THE LETTERS OF ONE LIST AGAINST THE OTHER.  WHEN THERE IS A */
     /* MATCH, THE PROGRAM SKIPS TO THE NEXT LETTER AND CHECKS IT AGAINST */
     /* THE LETTERS OF THE SECOND LIST.  IF NO MATCH IS FOUND, IT IS STORE*/
     /* D IN THE ARRAY P.  TO CHECK FOR THE POSSIBLITY THAT A MATCH MAY   */
     /* NOT OCCUR WHEN MATCHING EACH ELEMENT OF THE SECOND LIST AGAINST   */
     /* THE FIRST, THE SEARCH IS PERFORMED IN THE OPPOSITE DIRECTION      */
     /*                                                                           */
 17   KP=1;


 18   LOOP:DO I=1 TO N;            DO J=1 TO NE;
 20   IF X(L,I)=X(LL,J) THEN GO TO HELL;
 22  END;  P(KB,KP)=X(L,I);     KP=KP+1;
 25  HELL: END LOOP;
 26  NEVER:DO I=1 TO NE;       DO J=1 TO N;
 28   IF X(LL,I)=X(L,J) THEN GO TO HEAVEN;          END;
 31      KPP=KP-1;
 32       ZE=X(LL,I);
     /*                                                                           */
     /* THOSE LETTERS THAT HAVE NO MATCH IN EACH LIST ARE SORTED TO APPEAR*/
     /* IN A NICE MANNER         .    THESE ARE THE FINISHED PRODUCT.      */
     /*                                                                    */
 33    LAP:DO KK=1 TO KPP;
 34       IF ZE <P(KB,KK) THEN DO;
 36        TE=P(KB,KK);
 37        P(KB,KK)=ZE;
 38        ZE=TE;
 39      END;          END LAP;
 41       P(KB,KP)=ZE;       KP=KP+1;     HEAVEN:END NEVER;
    /*  COMPUTATIONS FOR THE THREE FACTOR INTERACTION TERMS.                  */
 44   KE=1;
 45   IF TRIP=0 THEN GO TO ZAP;
    /* CHECKS THE TWO TERM INTERACTIONS AGAINST A THIRD LIST FOR A MATCH */
 47   DO K=LL+1 TO MZ;
 48     KP=1;
 49   DO I=1 TO KP-1;      DO J=1 TO M(K);
    /* IF A MATCH OCCURS, SKIP TO THE NEXT LETTER IN THE LIST.  IF NO     */
    /* MATCH OCCURS, THEN ASSIGN THAT LETTER TO THE ARRAY S.            */
 51   IF P(KB,I)=X(K,J) THEN GO TO HER;
 53   END;
 54   S(KB,KE)=P(KB,I);
 55   KE=KE+1;
 56  HER:END;
```

```
/*                                                                    */
/* NEXT, WE CHECK FOR NO MATCH OF LETTERS IN THE REVERSE ORDER        */
/*                                                                    */
 REP:DO I=1 TO N(K);
  DO J=1 TO KP-1;
  IF X(K,I)=P(KB,J) THEN GO TO SHOW;
  END;
  S(KB,KE)=X(K,I);
  KE=KE+1;
 SHOW:END REP;
  NF(KB)=KE-1;
  KB=KB+1;
 END;
  KB=KP;
  DO I=1 TO KB-1;
  SW=1;
/* SORT THE LETTERS IN THE THIRD ORDER INTERACTION TERMS INTO ORDER.*/
 DO WHILE(SW¬=0);           SW=0;
  DO J=2 TO NF(I);
  IF S(I,J)<S(I,J-1) THEN DO;
  TE=S(I,J);           S(I,J)=S(I,J-1);         S(I,J-1)=TE;
  SW=1;
END;
  END;
 END;
/* PLACE THE THIRD ORDER INTERACTIONS BACK INTO THE ARRAY P.    */
  DO J=1 TO NF(I);
  P(I,J)=S(I,J);
  END;
  END;
  GO TO HARP;
  ZAP:NF(KB)=KP-1;
  KP=KP+1;
 HARP:END ALIAS;


/* END OF THE SUBROUTINE ALIAS.                                       */
/* READ IN MZ, THE NUMBER OF LISTS TO.BE COMBINED IN TWO AND THREE    */
/* INTERACTIONS.  NEXT READ IN THE LENGTH OF THE FIRST LIST AND THEN*/
/* READ IN THAT LIST INTO THE ARRAY A.  THIS IS FOLLOWED BY THE      */
/* NUMBER OF LETTERS IN THE NEW CODING SCHEME, WHICH IN TURN IS       */
/* FOLLOWED BY THE LIST FOR THE NEW CODINGS.   HIS IS REPEATED FOR    */
/* AS MANY AS INDICATED IN MZ.                                        */
GET EDIT(MZ) (COL(1),F(2));
DO I=1 TO MZ;
 GET EDIT(N,(A(I,J) DO J=1 TO N),NA,(BZ(J,I) DO J=1 TO NA))
   (X(1),F(1),X(1),(N) A(1),X(1),F(1),X(1),(NA) A(1));
   F(I)=N;
 NF(I)=NA;
  END;
  DO TRIP=0 TO 1;
  JJ=1;
 DO L=1 TO MZ;
  DO LL=1 TO MZ;
   KR=1;
 IF L>=LL THEN GO TO SEW;
/* CALL THE SUBROUTINE ALIAS TO COMPUTE THE TWO AND THREE WAY         */
/* INTERACTION TERMS.                                                 */
  CALL ALIAS(N(L),N(LL),N,NF,A,L,LL,MZ,P,KA,KR,TRIP);
 KC=KR-1;
DO I=1 TO KC;
/* STORE THE FINDINGS IN THE MATRIX R, CONCATENATING EACH LETTER TO */
/* FORM A NICE STRING OF CHARATERS TO BE OUTPUTTED.                  */
/* THIS IS FOR THE ORIGINAL OR OLD CODING SCHEME                    */
. R(JJ)='      ';
  IF NF(I)=1 THEN R(JJ)=P(I,1);
 ELSE IF NF(I)=2 THEN R(JJ)=P(I,1)||P(I,2);
  ELSE IF NF(I)=3 THEN R(JJ)=P(I,1)||P(I,2)||P(I,3);
 ELSE IF NF(I)=4 THEN R(JJ)=P(I,1)||P(I,2)||P(I,3)||P(I,4);
 ELSE IF NF(I)=5 THEN R(JJ)=P(I,1)||P(I,2)||P(I,3)||P(I,4)||P(I,5);
ELSE IF NF(I)=6 THEN R(JJ)=P(I,1)||P(I,2)||P(I,3)||P(I,4)||P(I,5)||
P(I,6);
```

```
   JJ=JJ+1;
   END;
 SEN:END;
   END;
   KK=1;
   DO I=1 TO MZ;    DO J=1 TO MZ;    IF I>=J THEN GO TO JIM;
/* THE SECOND ORDER INTERACTION TERMS FOR THE NEW CODING ARE COMPUTED*/
   IF TRIP=0 THEN DO;
  IF MM(I)=2&MM(J)=2 THEN Q(KK)=BZ(1,I)||BZ(2,I)||BZ(1,J)||BZ(2,J);
  ELSE IF MM(I)=1&MM(J)=2 THEN Q(KK)=BZ(1,I)||BZ(1,J)||BZ(2,J);
  ELSE IF MM(I)=2&MM(J)=1 THEN Q(KK)=BZ(1,I)||BZ(2,I)||BZ(1,J);
  ELSE Q(KK)=BZ(1,I)||BZ(1,J);
   KK=KK+1;
   GO TO JIM;
   END;
/* THE THIRD ORDER INTERACTION TERMS FOR NEW CODING SCHEME ARE        */
/* CREATED                                                             */
   DO K=1 TO MZ;
   IF K<=I|K<=J THEN GO TO KIM;
   IF MM(I)=2 THEN DO;
  PP=BZ(1,I)||BZ(2,I);
  IF MM(J)=2&MM(K)=2 THEN Q(KK)=PP||BZ(1,J)||BZ(2,J)||BZ(1,K)||BZ(2,K);
  ELSE IF MM(J)=1&MM(K)=2 THEN Q(KK)=PP||BZ(1,J)||BZ(1,K)||BZ(2,K);
  ELSE IF MM(J)=2&MM(K)=1 THEN Q(KK)=PP||BZ(1,J)||BZ(2,J)||BZ(1,K);
  ELSE Q(KK)=PP||BZ(1,J)||BZ(1,K);
  END;
  ELSE DO;
  IF MM(J)=2&MM(K)=2 THEN Q(KK)=BZ(1,I)||BZ(1,J)||BZ(2,J)||BZ(1,K)||
  BZ(2,K);
 ELSE IF MM(J)=1 & MM(K)=2 THEN Q(KK)=BZ(1,I)||BZ(1,J)||BZ(1,K)||
  BZ(2,K);
 ELSE IF MM(J)=2 & MM(K)=1 THEN Q(KK)=BZ(1,I)||BZ(1,J)||BZ(2,J)||
  BZ(1,K);
  ELSE Q(KK)=BZ(1,I)||BZ(1,J)||BZ(1,K);
  END;   KK=KK+1;   KIM:END;  JIM:END;    END;


  LQ=MZ*(MZ-1)/2;
   IF TRIP¬=0 THEN LQ=MZ*(MZ-1)*(MZ-2)/6;
/* SORT THE ALIASES FOR THE OLD CODING SCHEME INTO ASCENDING ORDER */
/* THIS IS DONE ONLY IF THE NUMBER OF ALIASES ARE LESS THAN 1000.  */
/* WITH MORE THAN  1000 THE COMPUTER TIME IS TOO COSTLY.  FOR THE    */
/* SITUATION WHERE THE NUMBER OF ALAISES ARE GREATER THAN 1000, THEY */
/* ARE OUTPUTTED TO AN EXTERNAL FILE ON DISK OR TAPE.  USING IBM SORT*/
/* ROUTINE, WHICH IS MUCH FASTER, THE ALIASES ARE SORTED FOR THE OLD */
/* CODING SCHEME.  THEN IN ANOTHER SHORT PROGRAM, CONTIN, THE SORTED */
/* ALIASES ARE READ BACK INTO THE COMPUTER AND OUTPUTTED IN NICE FORM*/
/* ALONG WITH THE NEW CODING SCHEME.                                 */
/* IF THE NUMBER OF ALIASES ARE LESS THAN 1000, THE PROCESS OR SORTIN*/
/*G AND OUTPUTTING ARE AUTOMATCI.                                    */
 IF LQ>1000 THEN GO TO LOOP;
   DO J=1 TO LQ;
   LOC(J)=J;     END;
   SW=1;          DO WHILE (SW¬=0);        SW=0;
  DO J=2 TO LQ;           L=LOC(J);             LL=LOC(J-1);
  IF R(L)<R(LL) THEN DO;    ITT=LOC(J);                LOC(J)=LOC(J-1);
  LOC(J-1)=ITT;    SW=1;        END;       END;        END;
   DO J=1 TO LQ;      KOC(J)=INDEX(R(J),' ')-1;      END;
   SW=1;    DO WHILE(SW¬=0);    SW=0;
   DO J=2 TO LQ;    L=LOC(J);    LL=LOC(J-1);
 IF KOC(L)<KOC(LL) THEN DO;    ITT=LOC(J);
  LOC(J)=LOC(J-1);     LOC(J-1)=ITT;          SW=1;
  END;                  END;                   END;
 PUT LIST(' '); PUT SKIP(4);
 IF TRIP=0 THEN GO TO MARS;    LK=1;
 PUT EDIT(R(LOC(1)),Q(LOC(1))) (COL(3),A(6),X(2),A(6));
  DO I=2 TO LQ; LK=LK+1; IF R(LOC(I))¬=R(LOC(I-1)) THEN DO;
 LK=1; PUT EDIT(R(LOC(I))) (SKIP(3),COL(3),A(6));    END;
 IF LK>15 THEN DO;  PUT EDIT(Q(LOC(I))) (COL(11),A(6));
 LK=1;  GO TO Y;  END;
```

```
PUT EDIT(Q(LOC(I))) (X(2),A(6));
Y:END;
 PUT SKIP(3);
 GO TO QQ;
LOCR:
     PUT EDIT(' THE DATA TO BE SORTED WAS TOO MUCH.  IT IS BEING WRITTEN
TO AN EXTERNAL FILE') (SKIP(4),COL(1),A);
 PUT EDIT(' TO RECOVER SORTED DATA, USE IBM SORT ROUTINE AND THEN THE
SHORT PROGRAM CONTIN TO OUTPUT FINAL ALIAS') (SKIP(3),COL(1),A);
DO J=1 TO LQ;
PUT FILE(GCOG) EDIT(R(J),KOC(J),Q(J)) (COL(1),A(6),X(2),F(4),X(2),A(6));
 END;
GO TO QQ;       NAMS:
NSP=NZ/2;   NPA=1;   LKK=1;   LLL=KIP*NSP;   NPE=3;
     DO KAK=1 TO LQ BY LLL;
PUT SKIP(4);  PUT EDIT(R(LOC(NPA))) (COL(3),A(6));
 NPA=NSP+NPA;
     DO I=NPA TO LQ BY NSP;
 LKK=LKK+1;      IF LKK>KIP THEN DO;
 NPA=I;      LKK=1;           GO TO SEA;
 END;
PUT EDIT(R(LOC(I))) (X(2),A(6));
 END;
SEA: NZ=KAK+NSP-1;                 PUT SKIP(3);
   DO J=KAK TO NZ;
KLL=1;     PUT EDIT(Q(LOC(J))) (COL(3),A(6));
     DO I=J+NSP TO LQ BY NSP;
KLL=KLL+1;  IF KLL>KIP THEN DO;  KLL=1;  GO TO LOC;
 END;
 PUT EDIT(Q(LOC(I))) (X(2),A(6));
          END;   LOO:END;        END;
QQ:END;
END ALPO;
```

FORMAT FOR ENTERING DATA CARDS

CARD 1

Columns 1 - 2	NLIST: number of lists
Column 3	Blank
Column 4	NLETT: number of letters in first string (list, old coding)
Column 5	Blank
Column 6	First letter of first string.
Column 7	Second letter of first string.
.	.
etc.	etc.
.	.
Column n	Last letter of first string.
Column n+1	Blank
Column n+2	NNEW: Number of letters in first string (list, new coding)
Column n+3	Blank
Column n+4	First letter of first list (new coding)
Column n+5	Second letter of first list (new coding)
.	. . . . . .
etc.	etc.
.	. . . . . .
Column n+m	Last letter of first list (new coding) if any.
Column n+m+1	Blank

FORMAT FOR ENTERING DATA CARDS (con't)

For the next string (second) or list, the specifications given in columns 4 through n+m+1 are repeated. This is repeated for as many lists as indicated in NLIST, columns 1 and 2.

An example of the data cards specification is given below for 32 lists.

CARD 1

```
Columns
         111111111122222222223333333333444444444455555555556666666666
123456789012345678901234567890123456789012345678901234567890123456789

32 6 ABCDEF 1 A 5 ABCDE 1 B 5 ABCDF 1 C 5 ABCEF 1 D 5 ABDEF 1 E 5 ACD

Columns

77777777778
01234567890

EF 1 F 4 AB
```

CARD 2

```
Columns

         111111111122222222223333333333444444444455555555556666666666
123456789012345678901234567890123456789012345678901234567890123456789

CD 1 G 4 ABCE 1 H 4 ABCF 1 I 4 ABDE 1 J 4 ABDF 1 K 4 ABEF 1 L 4 ACDE

Columns

77777777778
01234567890

1 M 4 ACDF
```

FORMAT FOR ENTERING DATA CARDS (con't)

CARD 3

Columns

```
         111111111122222222223333333333444444444455555555556666666666
123456789012345678901234567890123456789012345678901234567890123456789

1 N 4 ACEF 1 O 4 ADEF 1 P 3 ABC 1 Q 3 ABD 1 R 3 ABE 1 S 3 ABF 1 T 3 A
```

Columns

```
77777777778
01234567890

CD 1 U 3 AC
```

CARD 4

Columns

```
         111111111122222222223333333333444444444455555555556666666666
123456789012345678901234567890123456789012345678901234567890123456789

E 1 V 3 ACF 1 W 3 ADE 1 X 3 ADF 1 Y 3 AEF 1 Z 2 AB 2 A' 2 AC 2 B' 2 A
```

Columns

```
77777777778
01234567890


D 2 C' 2 AE
```

CARD 5

Columns

```
         111111111122222222223333333333444444444455555555556666666666
123456789012345678901234567890123456789012345678901234567890123456789

 2 D' 2 AF 2 D' 1 A 2 F'
```

FORMAT FOR ENTERING DATA CARDS (con't)

OLD CODING	NEW CODING
ABCDEF	A
ABCDE	B
ABCDF	C
ABCEF	D
ABDEF	E
ACDEF	F
ABCD	G
ABCE	H
ABCF	I
ABDE	J
ABDF	K
ABEF	L
ACDE	M
ACDF	N
ACEF	O
ADEF	P
ABC	Q
ABD	R
ABE	S
ABF	T
ACD	U
ACE	V

FORMAT FOR ENTERING DATA CARDS (con't)

OLD CODING	NEW CODING
ACF	W
ADE	X
ADF	Y
AEF	Z
AB	A'
AC	B'
AD	C'
AE	D'
AF	D'
A	F'

JOB CONTROL LANGUAGE (JCL) CARDS FOR IBM 360/91

```
//EII583CS  JOB  '060,4KI,ID=ALIES','HOWARD LEE'
// PASSWORD  xxxxxxxx
//STEP1   EXEC   PL1, RG=280K
//SYSIN DD *

     ( Source Program )
/*
//GO.GOOG DD DSN=EII583.HBL.ALIAS, VOL=SER=PUBLIC, UNIT=2314,
//        SPACE=(TRK,(10,5)),DCB=(RECFM=FB,LRECL=14,BLKSIZE=1400)
//GO.SYSIN DD *
     (data cards )
/*
//
```

APPENDIX V

PROBABILITY VALUES FOR CONSTRUCTING HALF-NORMAL GRIDS

The table below provides the probability values at which the first four largest effects would be plotted on grids for grids with from 63 to 8 ranks. For each set, the number in parentheses is the rank, R_s, where

$$R_x = .683\ Y + 0.5 \qquad (Y = \text{largest rank; also } N\text{-}1)$$

representing the estimated standard deviation for a Y-size grid (Daniel, 1959, p 322). The relationship between P, the probability value on normal probability paper, and P', the new probability values for the half-normal grid, is explained on page 85 in the text. Grids can be properly spaced by relating the original P values to their corresponding Z-values (where $\sigma = 1$) found in most normal distribution tables.

RANK	P'	P	RANK	P'	P	RANK	P'	P	RANK	P'	P
63	99.21	99.60	62	99.19	99.60	61	99.18	99.59	60	99.17	99.58
62	97.62	98.81	61	97.58	98.79	60	97.54	98.77	59	97.50	98.75
61	96.03	98.02	60	95.97	97.98	59	95.90	97.95	58	95.83	97.92
60	94.44	97.22	59	94.35	97.18	58	94.26	97.13	57	94.17	97.08
(44)			(43)			(42)			(41)		
59	99.15	99.58	58	99.14	99.57	57	99.12	99.56	56	99.11	99.55
58	97.46	98.73	57	97.41	98.71	56	97.37	98.68	55	97.32	98.66
57	95.76	97.88	56	95.69	97.84	55	95.61	97.81	54	95.54	97.77
56	94.07	97.03	55	93.97	96.98	54	93.86	96.93	53	93.75	96.88
(41)			(40)			(39)			(39)		
55	99.09	99.55	54	99.07	99.54	53	99.06	99.53	52	99.04	99.52
54	97.27	98.64	53	97.22	98.61	52	97.17	98.58	51	97.12	98.56
53	95.45	97.73	52	95.37	97.69	51	95.28	97.64	50	95.19	97.60
52	93.64	96.82	51	93.52	96.76	50	93.40	96.70	49	93.27	96.63
(38)			(37)			(37)			(36)		
51	99.02	99.51	50	99.00	99.50	49	98.98	99.49	48	98.96	99.48
50	97.06	98.53	49	97.00	98.50	48	96.94	98.47	47	96.88	98.44
49	95.10	97.55	48	95.00	97.50	47	94.90	97.45	46	94.79	97.40
48	93.14	96.57	47	93.00	96.50	46	92.86	96.43	45	92.71	96.35
(35)			(35)			(34)			(33)		
47	98.94	99.47	46	98.91	99.46	45	98.89	99.44	44	98.86	99.43
46	96.81	98.40	45	96.74	98.37	44	96.67	98.33	43	96.59	98.30
45	94.68	97.34	44	94.56	97.28	43	94.44	97.22	42	94.32	97.16
44	92.55	96.28	43	92.39	96.20	42	92.22	96.11	41	92.05	96.02
(33)			(32)			(31)			(31)		
43	98.84	99.42	42	98.81	99.40	41	98.78	99.39	40	98.75	99.38
42	96.51	98.26	41	96.43	98.24	40	96.34	98.17	39	96.25	98.12
41	94.19	97.09	40	94.05	97.02	39	93.90	96.95	38	93.75	96.88
40	91.86	95.93	39	91.67	95.83	38	91.46	95.73	37	91.25	95.62
(30)			(29)			(29)			(28)		

PROBABILITY VALUES FOR CONSTRUCTING HALF-NORMAL GRIDS (Continued)

RANK	P'	P	RANK	P'	P	RANK	P'	P	RANK	P'	P
39	98.72	99.36	38	98.68	99.34	37	98.65	99.32	36	98.61	99.31
38	96.15	98.08	37	96.05	98.03	36	95.95	97.97	35	95.83	97.92
37	93.59	96.79	36	93.42	96.71	35	93.24	96.62	34	93.06	96.53
36	91.03	95.51	35	90.79	95.39	34	90.54	95.27	33	90.28	95.14
(27)			(26)			(26)			(25)		
35	98.57	99.29	34	98.53	99.26	33	98.48	99.24	32	98.44	99.22
34	95.71	97.86	33	95.59	97.79	32	95.45	97.73	31	95.31	97.66
33	92.86	96.43	32	92.65	96.32	31	92.42	96.21	30	92.19	96.09
32	90.00	95.00	31	89.71	94.85	30	89.39	94.70	29	89.06	94.53
(24)			(24)			(23)			(22)		
31	98.39	99.19	30	98.33	99.17	29	98.28	99.14	28	98.21	99.11
30	95.16	97.58	29	95.00	97.50	28	94.83	97.41	27	94.64	97.32
29	91.94	95.97	28	91.67	95.83	27	91.38	95.69	26	91.07	95.54
28	88.71	94.35	27	88.33	94.17	26	87.93	93.97	25	87.50	93.75
(22)			(21)			(20)			(20)		
27	98.15	99.07	26	98.08	99.04	25	98.00	99.00	24	97.92	98.96
26	94.44	97.22	25	94.23	97.12	24	94.00	97.00	23	93.75	96.88
25	90.74	95.37	24	90.38	95.19	23	90.00	95.00	22	89.58	94.79
24	87.04	93.52	23	86.54	93.27	22	86.00	93.00	21	85.42	92.71
(19)			(18)			(18)			(17)		
23	97.83	98.91	22	97.73	98.86	21	97.62	98.81	20	97.50	98.75
22	93.48	96.74	21	93.18	96.59	20	92.86	96.43	19	92.50	96.25
21	89.13	94.57	20	88.64	94.32	19	88.10	94.05	18	87.50	93.75
20	84.78	92.39	19	84.09	92.05	18	83.33	91.67	17	82.50	91.25
(16)			(16)			(15)			(14)		
19	97.37	98.68	18	97.22	98.61	17	97.06	98.53	16	96.88	98.44
18	92.11	96.05	17	91.67	95.83	16	91.18	95.59	15	90.62	95.31
17	86.84	93.42	16	86.11	93.06	15	85.29	92.65	14	84.38	92.19
16	81.58	90.79	15	80.56	90.28	14	79.41	89.71	13	78.12	89.06
(13)			(13)			(12)			(11)		
15	96.67	98.33	14	96.43	98.21	13	96.15	98.08	12	95.83	97.92
14	90.00	95.00	13	89.29	94.64	12	88.46	94.23	11	87.50	93.75
13	83.33	91.67	12	82.14	91.07	11	80.77	90.38	10	79.17	89.58
12	76.67	88.33	11	75.00	87.50	10	73.08	86.54	9	70.83	85.42
(11)			(10)			(9)			(9)		
11	95.45	97.73	10	95.00	97.50	9	94.44	97.22	8	93.75	96.88
10	86.36	93.18	9	85.00	92.50	8	83.33	91.67	7	81.25	90.62
9	77.27	88.64	8	75.00	87.50	7	72.22	86.11	6	68.75	84.38
8	68.18	84.09	7	65.00	82.50	6	61.00	80.56	5	56.25	78.12
(8)			(7)			(7)			(6)		
7	92.86	96.43	6	91.67	95.83	5	90.00	95.00	4	87.50	93.75
6	78.57	89.29	5	75.00	87.50	4	70.00	85.00	3	62.50	81.25
5	64.29	82.14	4	58.33	79.17	3	50.00	75.00	2	37.50	68.75
4	50.00	75.00	3	41.67	70.83	2	30.00	65.00	1	12.50	56.25
(5)			(5)			(4)			(3)		

APPENDIX VI

DERIVATION OF COMBINED LINEAR AND CUBIC TREND-ADJUSTMENT EQUATIONS

Dr. Steve R. Webb

The following derivation parallels the ones used to obtain the linear and quadratic trend-correction equations described by Daniel and Wilcoxin (1966, pp 272-273).

1. Normal equations for ordered 2^P plans to correct for linear (L) and cubic (K) trends.

$$\ell\hat{L} + x\hat{X} + xX + yY + z\hat{Z} + \ldots\ldots\ldots\ldots = (L) \qquad (1.1)$$

$$k\hat{K} + x'\hat{X} + y'\hat{Y} + z'\hat{Z} + \ldots\ldots\ldots\ldots\ldots\ldots = (K) \qquad (1.2)$$

$$x'\hat{K} + x\hat{L} + N\hat{X} = (X) \qquad (1.3)$$

$$y'\hat{K} + y\hat{L} + N\hat{Y} = (Y) \qquad (1.4)$$

$$z'\hat{K} + z\hat{L} + n\hat{Z} = (Z) \qquad (1.5)$$

etc.

where $\ell = [LL]$, $x = [LX]$, $y = [LY]$, $z = [LZ]$

$k = [KK]$, $x' = [KX]$, $y' = [KY]$, $z' = [KZ]$

The meaning of the alternate symbols can be found in Table 15 in the text. $N = 2^P$ and (X), (Y), (Z) are the contrasts correlated with (L) and (K). A dot over a letter indicates it is an unknown term.

2. From equations (1.3), (1.4) and (1.5) we can obtain

$$N\hat{X} = (X) - x\hat{L} - x'\hat{K} \qquad (1.6)$$

$$N\hat{Y} = (Y) - y\hat{L} - y'\hat{K} \qquad (1.7)$$

$$N\hat{Z} = (Z) - z\hat{L} - z'\hat{K} \qquad (1.8)$$

etc.

3. Substituting these equations into (1.1) and (1.2) we obtain:

$$(N\ell - x^2 - y^2 - z^2 - \ldots)\hat{L} + (-xx' - yy' - \ldots)\hat{K} = N(L) - x(X) - y(Y) \ldots (1.9)$$

$$(-x'x - y'y \ldots)\hat{L} + (Nk - x'^2 - y'^2 - \ldots)\hat{K} = N(K) - x'(X) - y'(Y) \ldots\ldots (1.10)$$

With the solutions for L and K in terms of the observations and the design parameters, we can evaluate the regression coefficients directly from equations (1.6) to (1.8).

Equations 1.9 and 1.10 are written using the alternate symbols in Table 15, Equations IIIa and b, in the text.

APPENDIX VII
CALCULATING DETERMINANTS

Determinants for a 2 x 2 matrix as shown in this illustration are easy to calculate. For example, if the elements of the matrix were:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then the determinant of the matrix (indicated by the vertical lines), is:

$$D = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc)$$

when the a and d are sum of squares and b and c, sum of products in our application.

If there are three responses, then the matrices become larger to include the additional sum of products (e.g., between responses 1 and 2, 2 and 3, and 1 and 3. Thus, for three responses, the total matrix, by way of illustration, would be:

$$T = \begin{pmatrix} ss_{t1} & sp_{t12} & sp_{t13} \\ sp_{t12} & ss_{t2} & sp_{t23} \\ sp_{t13} & sp_{t23} & ss_{t3} \end{pmatrix}$$

a symmetrical matrix with the sum of squares for each response, 1, 2, and 3, on the diagonal, and the sum of products in the appropriate columns and rows off the diagonal. The determinant of a 3 x 3 matrix is:

$$D = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + dhc - gec - dbi - ahf$$

A computer would be used to calculate determinants for larger matrices.

APPENDIX VIII

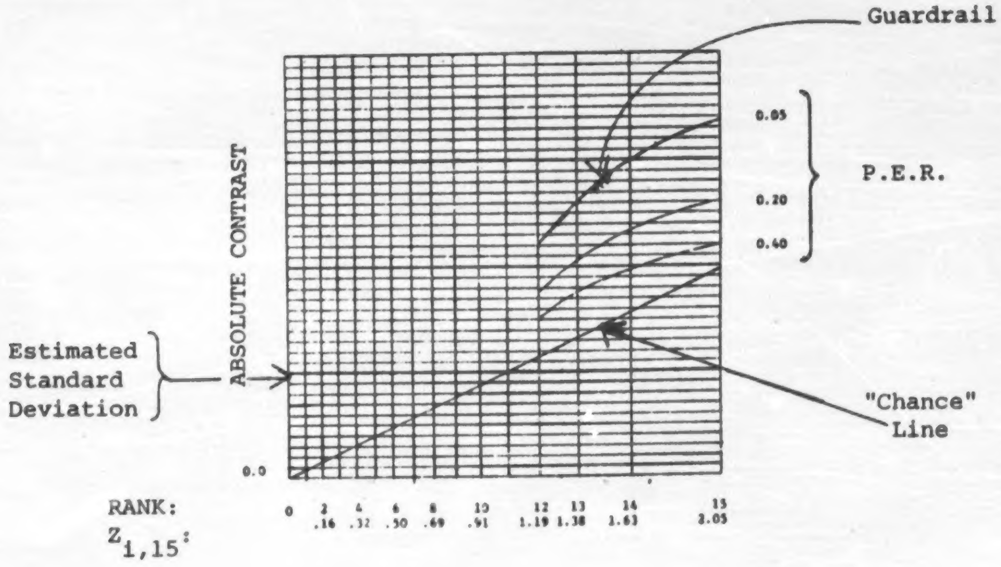
ZAHN'S GUARDRAILS FOR HALF-NORMAL PLOTS

Zahn (1975a) provides critical values for plotting guardrails for PER = $\alpha = 0.05$, 0.20, and 0.40 on the half-normal grids.

For version S, he provides them only for N = 15, assuming four real effects. This could be used if the results from a 2^{8-4}_{IV} screening design were plotted. The critical values, taken from Zahn's (1975a, p 197) Table 5, are:

α R	0.05	0.20	0.40
15	3.37	2.61	2.20
14	3.00	2.34	1.97
13	2.61	2.06	1.76
12	2.21	1.76	1.51

Unlike Daniel's, Zahn's guardrails will appear curved, as shown in this reproduction from his Figure 9 (p 198):



Appendix VIII (Continued)

For version X, Zahn (1975a, p 195) provides the critical values for N = 15, 31, 63, and 127. Taken from his Table 7, the critical values for N = 15, 31, and 63 are:

	R α	0.05	0.20	0.40
n = 15	15	3.230	2.470	2.066
	14	2.840	2.177	1.827
	13	2.427	1.866	1.574
	12	2.065	1.533	1.298
n = 31	31	3.351	2.730	2.372
	30	3.173	2.586	2.247
	29	2.992	2.439	2.121
	28	2.807	2.288	1.891
	27	2.615	2.133	1.857
n = 63	63	3.470	2.945	2.629
	62	3.384	2.872	2.564
	61	3.297	2.797	2.497
	60	3.209	2.722	2.431
	59	3.120	2.647	2.363
	58	3.030	2.570	2.295

DISTRIBUTION LIST

LCdr. James Ashburn, MSC, USN
NAMRL, Bldg. 1953
Pensacola, FL 32512

Dr. L. E. Banderet
SGDR-UE-CR
Dept. of the Army
U. S. Army Research Institute
of Environmental Medicine
Natick, Mass. 01760

Mr. Vernon E. Carter
Pilot Training Systems
Orgn 3750/62
Northrop Corp./Aircraft Div.
3901 W. Broadway
Hawthorne, CA 90250

Dr. Julien M. Christensen
Chairman, Dept. of Industrial Engr.
Wayne State University
Detroit, Michigan 48202

Dr. Chriss Clark
Honeywell, Inc. (MS R-2340)
2600 Ridgway Parkway
Minneapolis, Minn. 55413

Mr. James Duva (N-215)
Naval Training Equipment Ctr.
Orlando, FL 32813

Dr. Gordon A. Eckstrand
AFHRL/AS
Wright-Patterson AFB OH 45433

Mr. Ronald A. Erickson, Code 3175
Head, HF Branch, Weapons Devel. Dept.
U. S. Naval Weapons Center
China Lake, California 93555

Dr. Marshall J. Farr
ONR, Code 458
800 N. Quincy Street
Arlington, VA 22217

Terrence W. Faulkner
Health & Safety Laboratory
Bldg. 56, Kodak Park Division
Eastman Kodak Co.
Rochester, N. Y. 14650

Mr. Charles A. Gainer
Chief, Army Research Unit
Bldg. 502, P. O. Box 428
Ft. Rucker, Alabama 36360

Dr. Robert A. Goldbeck
Mail Station S-32
Western Development Laboratories Div.
Philco-Ford Corporation
3939 Fabian Way
Palo Alto, California 94303

James E. Goodson, CDR MSC USN
Head, Aerospace Psychology Dept.
Code 15, Naval Aerospace Med.
Research Lab.
Pensacola, Florida 32508

Dr. Tom Gray
AFHRL/FT
Williams AFB, AZ 85224

G. C. Helmstadter, Director
University Testing Services
Payne Hall, B302
Arizona State University
Tempe, Arizona 85281

Dr. Richard Jagacinski
Human Performance Center
330 Packard Road
Ann Arbor, Michigan 48104

Dr. Edgar M. Johnson
U. S. Army Research Institute for the
Behavioral and Social Sciences
1300 Wilson Blvd.
Arlington, VA 22209

AFHRL/ASM (Patricia A. Knoop)
Wright-Patterson AFB OH 45433

Distribution List (Continued)

Dr. Richard L. Krumm
P. O. Box 2706
Main Post Office
Washington, D. C. 20013

Mr. Robert G. Mills
6570th AMRL/HEB
Wright-Patterson AFB, OH 45433

Dr. Frederick A. Muckler (Code 311)
Prog. Dir., Design of Manned Systems
Navy Personnel R. & D. Center
San Diego, CA 92152

Dr. Wallace W. Prophet
Director, HumRRO Cent. Div.
400 Plaza Bldg.
Pensacola, FL 32505

Dr. Clyde R. Replogle
6750 AMRL/EME
Wright-Patterson AFB OH 45433

Charles V. Riche
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Marty Rockway
Technical Director
AFHRL/TT
Lowry AFB, CO 80230

Dr. Stanley N. Roscoe
Aviation Research Laboratory
Univ. of Ill. -- Willard Airport
Savoy, IL 61874

Dr. Mark Sanders
Department of Psychology
California State University
Northridge, CA 91324

Dr. Dennis E. Smith
Mathematical Statistician
Desmatics, Inc.
P. O. Box 863
State College, PA 16801

Dr. Margaret J. Smith
Naval Education and Training
Program Development Center
Ellyson Field
Pensacola, FL 32509

H. C. Strasel
Chief, ARI Field Unit
U. S. Army Research Institute
P. O. Box 2086
Ft. Benning, GA 31905

Dr. Martin A. Tolcott
Human Engineering Div., ONR
800 N. Quincy Street
Arlington, VA 22217

Dr. Donald A. Topmiller
AMRL/HES
Wright-Patterson AFB OH 45433

Dr. R. Young, Director
Human Resources Office, ARPA
1400 Wilson Blvd.
Arlington, VA 22209

HQ AFSC/DLS
Andrews AFB, MD 20334

HQ AFHRL/CC
Brooks, AFB, TX 78235

Director, Behavioral Sciences Dept.
USAF Academy
Colorado Springs, CO 80840

Flight Dynamics & Control Division
Mail Stop 152
NASA - Langley Research Center
Hampton, VA 23665
Attn: Gary P. Beasley

Department of the Air Force
Air Force Human Relations Lab. (AFSC)
Lackland AFB, TX 78236
Attn: Mark Nataupsky, Capt., USAF
Chief, Evaluation Section
Personnel Research Division

Distribution List (Continued)

Military Asst. for Human Resources
OAD (E&LS)
ODDR&E
Pentagon, Washington, D. C. 20330

Executive Editor
Psychological Abstracts
American Psychological Assn.
1200 17th St. N.W.
Washington, D.C. 20036

HQ USAF/RDPS
Washington, D.C. 20330

AFFDL/CC
Wright-Patterson AFB, OH 45433

Director
USAF Avionics Laboratory
Wright-Patterson AFB, OH 45433

AMD/RDH (Col. George C. Mohr)
Brooks AFB, TX 78235

Defense Documentation Center
Cameron Station
Alexandria, VA 22314

Education Research Information Center
Processing & Reference Facility
4833 Rugby Ave., Suite 303
Bethesda, MD 20014

NASA - Scientific & Technical
Information Facility
P. O. Box 8757
B.W.I. Airport, Maryland 21240

National Technical Information
Services (NTIS)
Operations Division
5285 Port Royal Road
Springfield, VA 22151

Elizabeth Lage Roscoe
2117 Bristol Road
Champaign, Illinois 61820

Dr. J. Robert Newman
Department of Psychology
California State College
Long Beach, CA 90840

Mr. R. S. Easterby
Applied Psychology Department
The University of Aston
College House, Gosta Green
Birmingham B47ET England

Dr. J. Peter Kincaid
Martin Marietta Aerospace
Orlando Division
P. O. Box 5837
Orlando, FL 32805

Dr. W. G. Matheny
Life Sciences Inc.
227 Loop 820 Northeast
Hurst, TX 76053

Dr. Thomas J. Triggs
Department of Psychology
Monash University
Clayton, Victoria 3168
Australia